



Excel GridWay

Plugin para el acceso transparente a infraestructuras grid desde Excel

Autores:

Jorge Carretero Cruzado

Daniel Salmerón Majadas

Laura Tejedor de la Fuente

Director: Ignacio Martín Llorente

Proyecto de Sistemas Informáticos 2006/2007

Facultad de Informática

Universidad Complutense de Madrid

Índice

0 RESUMEN DEL PROYECTO.....	8
0.1 Resumen.....	8
0.2 Abstract	8
0.3 Autorización	9
0.4 Lista de palabras clave para búsquedas bibliográficas.....	10
1 RESUMEN DE LA MEMORIA	11
Las redes grid	11
Globus Toolkit.....	11
Metaplanificador Gridway	12
Necesidades actuales	14
Las hojas de cálculo como herramienta de trabajo.....	14
Soluciones existentes	15
Excel Connector.....	15
Platform Symphony adapter for MS Excel	15
ActiveSheets.....	15
ExcelGrid.....	16
Novedades y mejoras que aportamos	16
El proyecto Excel Gridway	17
Conexión a una red Grid.....	18
Lanzamiento de un trabajo para su ejecución.....	18
Monitorizar un trabajo.....	19
Visualización de resultados de un trabajo	20
2 LAS REDES GRID.....	22
2.1. Introducción	22
¿Qué es Grid?	22
Motivación	22
Arquitectura	23
Avance respecto a la World Wide Web.....	24
Aplicaciones	24
2.2. MiddleWare (Globus)	26
¿Qué cubre Globus Toolkit?	27
“Connectivity Layer” o capa de conectividad, que se encarga de:	27
Gestión de servicios (WSRF)	27
“Resource Layer” o capa de recursos, que se encarga de:	28
“Collective Layer” o capa colectiva, que se encarga de:	28
Servicios básicos	28
Sus principales componentes se muestran en la siguiente figura:.....	28
Utilización del middleware Globus Toolkit	30
2.3 Metaplanificación (GridWay).....	31
3 PANORAMA ACTUAL	34
3.1 Necesidades actuales.....	34
Posibilidades de Grid en el entorno actual	34
Las hojas de cálculo como soporte de trabajo.....	35
3.2 Soluciones existentes.....	36

Excel Connector.....	36
Platform Symphony adapter for MS Excel	37
ActiveSheets.....	38
ExcelGrid.....	38
3.3 Novedades y mejoras que aportamos.....	42
4 EL PROYECTO PLUGIN EXCEL GRID	46
4.1 Planteamiento del proyecto.....	46
4.2 Casos de uso del sistema	46
Casos de uso:	46
4.2.1 Caso de uso “Conexión a red Grid”	47
4.2.2 Caso de uso “Lanzar un trabajo a ejecución”	48
4.2.3 Caso de uso “Monitorizar un trabajo”	49
4.2.4 Caso de uso “Cancelar un trabajo”	51
4.2.5 Caso de uso “Ver resultados de un trabajo”	52
4.3 Diseño del plugin Excel Grid	53
4.3.1 Restricciones y requisitos del sistema	54
4.3.1.1 Descripción general.....	54
4.3.1.2 Contexto de uso	54
Perfiles de usuario o participantes	54
Usos del software.....	54
Actores	55
4.3.1.3 Requisitos funcionales.....	55
Conexión a una red Grid.....	55
Lanzamiento de un trabajo para su ejecución.....	55
Monitorizar un trabajo.....	57
Visualización de resultados de un trabajo.....	57
4.3.1.4 Requisitos estéticos.....	57
La interfaz	57
4.3.1.5 Requisitos de utilidad	58
Facilidad de uso y aprendizaje	58
Personalización e internacionalización de los requisitos.....	58
Requisitos de accesibilidad	58
4.3.1.6 Requisitos operacionales.....	58
4.3.1.7 Requisitos de seguridad.....	59
4.3.1.8 Persistencia de datos.....	59
4.3.1.9 Restricciones y limitaciones.....	60
Requisitos Hardware	60
Requisitos Software	60
4.4 Entorno de implementación.....	61
4.4.1 Tecnologías y herramientas de desarrollo.....	61
4.4.1.1 Tecnología SSH	61
4.4.1.2 Tecnología Grid.....	61
4.4.1.3 Microsoft Visual Studio .NET	62
4.4.2 Medio físico.....	63
4.4.2 Software implicado en el desarrollo.....	63
5 CONCLUSIONES Y TRABAJO FUTURO	64
5.1 Conclusiones	64
5.1.1 Situación final del proyecto	64
5.1.2 Trabajo realizado y conocimientos adquiridos	64
5.2 Trabajo futuro.....	65
5.2.1 Ampliar la aplicación para que sirva para otro tipo metaplanificadores ..	65
5.2.2 Crear un modo interactivo para la aplicación	65
Apéndice A - Manual de usuario de Plugin Excel Grid	66

A.1 Introducción	66
A.2 Plugin Excel Grid.....	66
A.2.1 Instalación y ejecución de Plugin Excel Grid	66
Para instalar la aplicación.....	67
Para ejecutar la aplicación	70
Desde el escritorio	70
Desde el menú inicio	71
A.2.2 Requerimientos del sistema	72
A.2.3 Elementos del Plugin Excel Grid	72
A.3. Tutorial.....	74
A.3.1 Configuración de la conexión a la red Grid:	74
□ Host	75
□ User	75
□ Password	75
□ Grid Passphrase	75
□ Prompt	75
A.3.2 Crear y lanzar un nuevo trabajo a la red Grid:	76
Job Name	76
Executable	76
Executable Path.....	77
Executable File.....	77
Input Cell Range	77
A.3.3 Monitorizar un trabajo:.....	79
A.3.4 Cancelar un trabajo:.....	80
A.3.5 Visualizar los resultados de un trabajo:	81
Apéndice B - HowTo Inglés.....	83
B.1 Plugin Excel.....	83
OnConnection	83
OnDisconnection	84
OnAddInsUpdate	84
OnStartupComplete	84
OnBeginShutdown	84
O_cmd_Click.....	84
B.2 Plugin Excel Grid Runner	84
GSettings.....	85
GTrabajos.....	85
Mappings	85
mainClass	85
frmGClient.vb	86
B.3 Gridway Bridge	87
Introduction	87
1. Data configuration	87
2. Grid commands to operate with grid.....	89
3. Utilidades de apoyo internas de la librería	91
Apéndice C - HowTo Español	92
C.1 Plugin_excel	92
OnConnection	92
OnDisconnection	93
OnAddInsUpdate	93
OnStartupComplete	93
OnBeginShutdown	93
O_cmd_Click.....	93
C.2 Plugin_excel_grid_runner	94

GSettings	94
GTrabajos	94
Mappings	95
mainClass	95
frmGClient.vb	95
C.3 Plugin ExcelGridway.....	96
Introducción	96
1. Datos de configuración	97
2. Comandos de operación con grid	98
3. Utilidades de apoyo internas de la librería	100
Apéndice D - Bibliografía	102

0 RESUMEN DEL PROYECTO

0.1 Resumen

Mientras que la Web es un servicio para compartir información sobre Internet, el Grid es un servicio para compartir potencia computacional y capacidad para el almacenamiento de datos sobre Internet. El Grid va más allá de una simple comunicación entre computadores, y pretende convertir la red global de computadores en un gran recurso computacional.

Hoy, el Grid es un “trabajo en progreso”, con la tecnología subyacente aún en fase de prototipado, y está siendo desarrollado por cientos de investigadores e ingenieros de software a lo largo del mundo.

El Grid está atrayendo mucho interés debido a su futuro, que aunque incierto, es potencialmente revolucionario. Su interés viene no sólo de expertos en informática, sino de científicos, hombres de negocios, periodistas, y navegantes de la Web. Así, este proyecto, Excel Gridway es de gran interés principalmente para usuarios de Microsoft Excel.

Excel Gridway es un plugin o extensión para Microsoft Excel que permite enviar trabajos de dicha aplicación, que resulten muy pesados o que consuman mucha potencia computacional, a la red Grid para su ejecución y devolviendo los resultados sobre la misma hoja de cálculo. De esta manera, se libera a la máquina local de una excesiva carga de trabajo, y es totalmente transparente al usuario.

Así, este plugin supone una gran mejora para la aplicación Microsoft Excel, y una gran ayuda para sus usuarios que necesiten hacer trabajos muy pesados que requieran muchos recursos computacionales.

0.2 Abstract

Whereas the Web is a service for sharing information over the Internet, the Grid is a service for sharing computer power and data storage capacity over the Internet. The Grid goes beyond simple communication between computers, and aims ultimately to turn the global network of computers into one vast computational resource.

Today, the Grid is a "work in progress", with the underlying technology still in a prototype phase, and being developed by hundreds of researchers and software engineers around the world.

The Grid is attracting a lot of interest because its future, even if still uncertain, is potentially revolutionary. Its interest comes not only from experts in computer science, but from scientists, businessmen, journalists and, the browsers of the web. So, the interest of this project Excel Gridway, is mainly for users of Microsoft Excel.

Excel Gridway, is a plugin for Microsoft Excel that allows sending heavy and hard jobs of the application to the Grid for its execution and return the results on the worksheet. In this way, the local computer is released of a excessive load of work, and it is totally transparent to the user.

So, this plugin involves a great improvement for the Microsoft Excel application, and a great help for the users that need to do heavy jobs that require a lot of computational resources.

0.3 Autorización

Se Autoriza a la Universidad Complutense a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores, tanto la propia memoria, como el código, la documentación y/o el prototipo desarrollado:

Fdo. Jorge Carretero Cruzado

Fdo. Daniel Salmerón Majadas

Fdo. Laura Tejedor de la Fuente

En Madrid, a 4 de Julio de 2007.

0.4 Lista de palabras clave para búsquedas bibliográficas

Plugin, Excel, Grid, Globus, Gridway, template, bulk o array de tareas, proxy, SSH.

1 RESUMEN DE LA MEMORIA

Las redes grid

El proyecto hace uso de una tecnología emergente y que está causando gran impacto por su potencial, esta tecnología es el grid.

Podemos decir que llamamos grid al sistema de computación distribuido que permite compartir recursos no centrados geográficamente para resolver problemas de gran escala, de forma uniforme, transparente, segura, eficiente y fiable.

El proyecto utiliza esta tecnología de manera que un usuario de Excel, con necesidades importantes de cálculo, que su máquina no pueda realizar en un tiempo razonable, pueda delegar ese trabajo en la red grid, que recibirá las tareas y devolverá al usuario sus resultados en un espacio de tiempo mínimo.

De hecho podemos ver el grid como una tecnología análoga a las redes de suministro eléctrico, donde se utiliza un servicio cuando se necesita.

En definitiva la potencia que ofrecen multitud de computadores conectados en red usando grid es prácticamente ilimitada, además de que ofrece una perfecta integración de sistemas y dispositivos heterogéneos, por lo que las conexiones entre diferentes máquinas no generarán ningún problema.

Globus Toolkit

Existen diversas implementaciones de red grid, pero surge con fuerza el proyecto Globus y se convierte en un estándar de facto.

Globus no es una herramienta de usuario ni una aplicación, es un conjunto de librerías, servicios y API's para construir Grids. Es un software de tipo Middleware destinado a desarrolladores, que cuenta con documentación precisa para construir una red grid.

Globus proporciona seguridad, servicios básicos, librerías para escribir cualquier servicio web que un desarrollador desee, y algunas utilidades finales que facilitan enormemente las tareas como por ejemplo el protocolo gridftp, utilizado para envío de ficheros entre hosts del grid.

El proyecto utiliza un grid montado con Globus, a través del metaplanificador GridWay.

Metaplanificador Gridway

Grids ofrece un dramático crecimiento en el número de recursos de procesamiento y almacenamiento disponibles que pueden ser distribuidos a las aplicaciones. Sin embargo, el envío eficiente de trabajos y su gestión continúan siendo difícilmente accesibles a los científicos e ingenieros comunes debido a su naturaleza dinámica y compleja. La meta del proyecto Gridway es investigar y desarrollar la tecnología requerida para realizar automáticamente todos los pasos de envío de trabajos y también proporcionar los mecanismos de tiempo de ejecución requeridos para adaptar dinámicamente la ejecución de la aplicación.

El marco de trabajo Gridway ha sido desarrollado para reducir el vacío entre el middleware Grid y los desarrolladores de aplicaciones.

El marco de trabajo Gridway es un componente para la metaplanificación en el ecosistema Grid, previsto para usuarios finales y desarrolladores de aplicaciones Grid. Es un gestor de carga que realiza la gestión de la ejecución de trabajos y la repartición de recursos en un Grid formado por distintas plataformas de computación dirigidas por servicios Globus. Gridway permite desatendidos, confiables y ejecución eficiente de trabajos simples, complejos o en array en grids heterogéneos y dinámicos. Gridway realiza toda la planificación y envíos de trabajos de forma transparente al usuario final y adapta la ejecución de trabajos para cambiar las condiciones grid proporcionando mecanismos de recuperación de fallos, planificación dinámica. Gridway en Globus permite desacoplar las aplicaciones y los sistemas de gestión local subyacentes.

Gridway es un metaplanificador que permite una compartición de recursos computacionales (clusters, servidores, supercomputadoras...) de manera segura, eficiente y a larga escala, gestionado por diferentes sistemas LRM (Local Resource Management), como PBS, SGE, LSF, Condor..., con una organización simple o dispersada a través de varios dominios administrativos.

Gridway es un proyecto Globus, que cuenta con la filosofía Globus y con las directrices para el desarrollo colaborativo y contribuciones tanto de individuales como de corporaciones a lo largo del mundo.

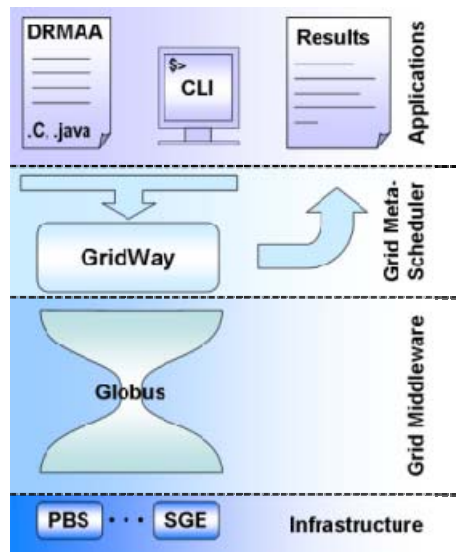
Existe un número de sistemas de gestión y planificación de la carga de trabajo tanto comerciales como de libre distribución disponibles hoy en día, cada uno adecuado para diferentes infraestructuras de computadores subyacentes y perfiles de ejecución. Gridway destaca entre otros sistemas metaplanificadores porque ha sido diseñado específicamente para trabajar por encima de los servicios de Globus, ofreciendo la más alta funcionalidad, calidad de servicio y confiabilidad en esta clase de infraestructuras.

Cuenta con muchas ventajas para:

- Directores de infraestructuras y proyectos: Gridway es un proyecto comunitario de libre distribución, que cuenta con la filosofía Globus y las directrices para el desarrollo colaborativo y no necesita de nuevos servicios, sino que trabaja con los mismos que Globus.
- Integradores de sistemas: Gridway es altamente modular, permitiendo la adaptación a diferentes infraestructuras grid y soporta varios estándares OGF. También es fácilmente extensible y podría ser utilizado o extendido por arquitecturas más complejas que implementen los acuerdos a nivel de servicio.
- Jefes de sistemas: Gridway proporciona un marco de trabajo de planificación similar al encontrado en los sistemas de gestión de recursos locales (LRM), permitiendo llevar la cuenta de los recursos y definiendo políticas de planificación.
- Desarrolladores de aplicaciones: Gridway implementa el API estándar DRMAA (C y Java), asegurando la compatibilidad de aplicaciones con sistemas LRM que implementan el estándar como SGE, Condor, Torque, ...
- Usuarios finales: Gridway permite a los usuarios finales enviar, monitorizar, sincronizar y controlar trabajos de forma totalmente transparente y similar los sistemas de gestión de recursos locales, que podrían ser descritos usando el estándar OGF JSDL. El planificador les facilita el trabajo ya que puede gestionar diferentes situaciones de fallo de forma automática.

Con Gridway, una infraestructura Grid puede ser explotada y gestionada de la misma manera que un cluster local.

Su arquitectura se detalla en la siguiente figura:



Necesidades actuales

En la actualidad no basta con utilizar un ordenador para dar solución a las necesidades diarias, se busca una progresiva mejora para potenciar los resultados obtenidos. Las máquinas son cada día más potentes pero esto en ocasiones no es suficiente para los objetivos que buscamos. Para obtener una mayor eficiencia se utilizan las redes gris, las cuales mediante la partición y distribución de tareas entre distintos equipos se obtienen una mejora en el procesamiento que se hace difícil de alcanzar a través de una única máquina. Además presenta una clara ventaja sobre el aspecto económico puesto que ofrece la posibilidad de contar con una mayor potencia de procesamiento sin necesidad de gastar grandes cantidades en equipos que estén a la altura del rendimiento esperado.

Hoy en día los principales impulsores de la tecnología GRID son las universidades y los centros de investigación y desarrollo. Pero los avances en el GRID no se quedan ahí y determinados sectores empresariales ya han empezado a ponerlos en práctica para obtener un mejor rendimiento en sus negocios.

Las hojas de cálculo como herramienta de trabajo

Son un instrumento de trabajo multifuncional que gozan de gran aceptación dentro de diversos campos, y que gracias a su sencillez de uso y su amplitud de recursos hacen que sean usadas para muy diversos fines, desde llevar un pequeño control de la contabilidad a nivel personal hasta grandes proyectos a nivel empresarial. Aparecieron en 1979 de la mano de VisiCalc y en aquella época eran consideradas como aplicaciones de alto nivel. No se centran en lo que puede conseguir la herramienta sino en como pueden ayudar al usuario a obtener mejores soluciones a sus problemas.

Una de las funciones que aporta MS Excel es la de trabajar como una base de datos, permitiendo un manejo fluido de grandes cantidades información. Combinando la variedad de las funciones disponibles y esta capacidad de almacenamiento, obtenemos una herramienta potente y versátil. En cada nueva versión de MS Excel que aparece en el mercado su capacidad de cálculo y de gestión de información aumenta. Es por esto que se desarrollan diversidad de complementos para MS Excel que le ayudan a aligerar las cargas de trabajo de muy diversas maneras (entre ellas la distribución de las tareas entre distintas entidades, el caso que aquí se plantea).

Microsoft provee a los programadores que quieran trabajar con Excel de un modelo de objetos que puede ser usado para controlarlos de una forma programada. Este modelo consta de un conjunto de clases y métodos que son los homólogos de sus correspondientes representaciones lógicas en Excel

Soluciones existentes

Excel Connector

Esta es una ampliación de las funcionalidades de Excel que trabaja directamente con InnerGrid, una herramienta enfocada a la utilización del GRID por parte del mundo empresarial. Trabaja de forma completamente transparente al usuario final, integrándose en el interfaz visual de Excel sin la necesidad de cambiar el método de trabajo que se aplicaba antes de utilizar esta extensión.

Platform Symphony adapter for MS Excel

Este software está enfocado hacia un mercado económico, para una toma de decisiones más rápida y efectiva a la hora de la concesión de créditos o el estudio de riesgos de mercado. Proporciona una mejora lineal de las simulaciones y garantiza una entrega de los resultados solicitados en el tiempo previsto; esto es así gracias a la combinación de la ejecución local y remota de las tareas. Trabaja con XLL's, (ficheros que definen la forma en la que los documentos XML nativos de Excel deben conectarse entre si), un modo estándar para representar enlaces entre recursos; la gestión de estos es un aspecto clave de esta aplicación mediante un acceso centralizado y una actualización continuada.

ActiveSheets

Se sustenta en otras dos aplicaciones desarrolladas por el mismo equipo denominadas Nimrod y EnFunzion. La máquina desde la que se invoca el trabajo se convierte en la máquina “raíz” siendo esta la que controla el experimento. El sistema, con la información de la que dispone, lanza el trabajo a la primera máquina disponible; cada nodo de trabajo puede disponer de un tipo de tecnología distinta. Las funciones de tratamiento de la información están

implementadas con Visual Basic for Applications (VBA) como stubs (un pequeño ejecutable que acapara parte del control al crear un trabajo), mediante la aplicación de automatización OLE. Esta aplicación presenta actualmente la posibilidad de que el usuario decida que funciones deben de ser ejecutadas en paralelo en la máquina remota.

ExcelGrid

Se trata de un sistema middleware a nivel de usuario. Colabora con el programa extendiéndolo para participar en redes corporativas (Alchemy, framework destinado al trabajo de redes GRID basado en Windows e implementado con tecnología Microsoft .NET) o globales (Gridbus, framework de trabajo multiplataforma que proporciona ciertos servicios que permiten la ejecución de tareas en diversos sistemas de middleware que trabajan a bajo nivel). Puede ejecutar aplicaciones externas a Excel que toman como parámetros los valores contenidos en las celdas seleccionadas de la hoja. Tiene dos componentes principales: Add-in y Runner. El primero es el encargado de la creación del acceso al plug-in desde Excel, el frontal de comunicación con el modelo de aplicación de Excel, mientras que el segundo es el GUI con el que interactúa el usuario desde donde se pueden configurar las opciones disponibles, almacenarlas, enviar trabajos y recibir información acerca de estos. La comunicación con el middleware se realiza mediante una conexión directa entre sockets y un protocolo específico para la especificación de los trabajos y sus parámetros. Utiliza tecnología COM a través de la característica COM Interop para trabajar con .NET.

Cada proceso necesario se envía al ordenador remoto en forma de trabajo formado por un ejecutable junto con los parámetros de entrada y el cliente grid permanece a la escucha de posibles eventos de este trabajo.

Novedades y mejoras que aportamos

Trabajando de manera totalmente transparente al usuario dedica de manera exclusiva sus recursos a la aplicación de estos sobre Microsoft Excel, mientras que otras extensiones no poseían esa dedicación absoluta. La principal diferencia es que nuestro plug-in trabaja con Gridway como cliente grid, un sistema gestor de carga y meta-planificador diseñado para trabajar exclusivamente sobre servicios provistos por GLOBUS.

A la hora de comunicarnos con el cliente GRID utilizamos un protocolo SSH a través de una librería especializada que incorporamos a nuestro proyecto (Sharp SSH). La ventaja que aporta SSH a nuestro desarrollo es la de proporcionar una comunicación mucho más segura para la transmisión de datos. Además la comunicación es directa entre el equipo que solicita la ejecución del trabajo y el cliente GRID, sin intermediación de ningún otro software. El plug-in está formado por una única entidad que trabaja directamente entre MS Excel y el

cliente grid, sin necesidad de terceros programas que gestionen conexiones ni manejen los datos que se envían.

Otro punto que aporta cierta flexibilidad a la aplicación es la posibilidad de elegir la localización del ejecutable que va a realizar los cálculos solicitados. Este ejecutable puede estar en el equipo local del usuario y ser enviado junto con los parámetros necesarios para su ejecución o bien estar alojado en los equipos remotos a los que se va a transmitir las entradas. Además el modo de gestionar los resultados obtenidos de los trabajos es diferente, no se aplica un método automático que los recoja, es el usuario el que decide cuando insertarlos en la hoja de cálculo. Quizá plantee en cierta medida la necesidad de una mayor participación por parte del usuario pero asegura un mayor control por parte de este. De esta manera se pueden gestionar de la forma deseada los resultados obtenidos, descartándolos y eliminándolos directamente del equipo remoto si finalmente no nos resultan necesarios. Gracias a esta gestión de la información evitamos contar con un número masivo de comunicaciones, al mismo tiempo que ganamos en seguridad.

El proyecto Excel Gridway

Excel Gridway es un plugin para la aplicación Microsoft Excel que permite enviar trabajos a una red Grid para su ejecución y depositar el resultado en la propia hoja de cálculo. De esta manera, se libera de una excesiva carga de la ejecución a la máquina local de forma totalmente transparente al usuario, y permite a los usuarios de Microsoft Excel que necesiten realizar cálculos muy complejos o que requieran mucha potencia de ejecución o recursos sin colapsar su máquina local, y de forma rápida y cómoda a través de dicha red.

Para ejecutar las funciones en remoto, el plugin se conecta a una red grid que es la que lleva a cabo la ejecución de las tareas y devuelve finalmente los resultados. Concretamente, es la red Gridway la que se utiliza en dicho plugin.

Para acceder a la red grid, se necesita autenticación, y por ello desde el mismo plugin, se activa dicha conexión para poder utilizar las funcionalidades que este ofrece, a través de algunos parámetros de configuración.

Cuando un usuario se conecta a la red grid tendrá las siguientes posibilidades:

- Envío de trabajos
- Monitorización de trabajos
- Cancelación de trabajos

- Obtención de resultados

Conexión a una red Grid

Para todas las operaciones del plugin Excel Grid, será necesario establecer previamente una conexión a una red Grid.

Para ello, el usuario rellenará los campos del formulario correspondientes a la configuración de la red, concretamente en la pestaña Settings del plugin.

Estos datos son el nombre del host remoto, el nombre de usuario, la contraseña de usuario, la contraseña del Proxy, y el Prompt de la máquina remota.

De esta manera, cada vez que se realice una tarea, como el envío de un trabajo, la cancelación de un trabajo, la monitorización de un trabajo o su visualización de resultados, el sistema establecerá previamente dicha conexión a partir de estos datos de configuración.

La forma de realizar la conexión será por una parte la inicialización de la sesión SSH, que será la técnica utilizada para la comunicación con la máquina remota desde Windows. Para ello se utilizarán los siguientes datos de configuración: el nombre del host remoto, el nombre de usuario, la contraseña de usuario, y el Prompt de la máquina remota.

Posteriormente se inicializará el Proxy, para lo cual se utilizará la clave del Proxy introducida por el usuario.

Una vez realizadas estas acciones, el sistema estará preparado para cualquier tipo de comunicación con la red grid.

Lanzamiento de un trabajo para su ejecución

La principal finalidad del proyecto, es la posibilidad de lanzar trabajos desde la aplicación Microsoft Excel a una red Grid, para su ejecución alguna máquina remota de esa red.

Para ello, el usuario rellenará los campos del formulario correspondientes a la configuración del trabajo, concretamente en la pestaña Submit Job del plugin.

Estos datos son:

- El nombre para el trabajo

- El ejecutable: Se permitirá al usuario elegir entre un ejecutable remoto (que se encontrará en la máquina destino), del que se especificará el nombre, o bien, un ejecutable local, del que se especificará la ruta completa para su envío a la red grid.
- El rango de celdas de entrada: El usuario seleccionará el rango de celdas de entrada sobre cuyos valores se aplicará la función.

Para el lanzamiento del trabajo, primero se formarán los ficheros de entrada “input” que constituirán los parámetros de entrada de la función y el fichero plantilla o “template”.

Estos ficheros se almacenarán en un directorio que llevará el nombre del trabajo especificado por el usuario, y que se enviará a la red grid.

Se podrán tener varios parámetros de entrada para un mismo ejecutable. Así, el número de columnas especificadas en el rango de celdas de entrada, será el número de parámetros de entrada para el ejecutable, e irá en un mismo fichero “input”.

Así mismo, el número de filas existentes también jugará un papel importante, ya que el lanzamiento del trabajo será un envío de un “bulk” o array de tareas, de manera que cada fila será una de ellas, y por tanto un fichero “input” cuya extensión será el número de la tarea dentro del array.

Una vez formados estos ficheros “input”, se formará el fichero “template”, en el cual se especificará el nombre del ejecutable, el formato del nombre de los ficheros de entrada “input”, el de los ficheros de salida “output”, y el de los ficheros de error “err”.

El siguiente paso es realizar la conexión a la red grid, y llamar a la función gwsbmit especificando el número de tareas del array y el directorio donde se encuentran tanto el template, como los ficheros de entrada “input”.

Una vez realizado el lanzamiento del trabajo se le devolverá al usuario un mensaje de información con el número de ID que tiene su trabajo en la red grid, para que le sea posible la posterior monitorización y visualización de resultados.

Monitorizar un trabajo

Otra de las funcionalidades principales de la aplicación, es la posibilidad de monitorizar los trabajos enviados y poder controlar su estado en todo momento.

Para ello, existirá una tabla de trabajos, en la que se muestren todos los trabajos enviados y su estado. El usuario podrá actualizar dicha tabla en cualquier momento para seguir la evolución de sus trabajos.

Así mismo, si desea una información más precisa del estado de algún trabajo, lo podrá seleccionar en la tabla y tendrá una opción de visualizar todos los detalles de dicho trabajo.

Para llevar a cabo esta monitorización, se realiza una conexión a la red grid, y se llama a la función gwps, que devuelve el historial de trabajos lanzados, y con esos datos será con los que se rellene la tabla de trabajos.

Visualización de resultados de un trabajo

Una vez que alguno de los trabajos se encuentra en estado zombie o completado, es posible visualizar los resultados de la ejecución en la hoja de cálculo de Microsoft Excel.

Para ello, el usuario escogerá el rango de celdas de salida en la hoja de cálculo donde se colocarán los resultados.

Se comprobará que el espacio de celdas elegido corresponde con el número de tareas del array del trabajo, y una vez que todo es correcto se actualizarán dichos resultados.

Para esta actualización, se hará previamente una conexión a la red Grid, y se traerán los ficheros “output” o de resultado, así como los “err” o ficheros de error a la máquina local. Posteriormente se abrirán dichos ficheros y se sacarán los resultados a las celdas.

Aparte de estas funcionalidades principales del plugin, también cabe destacar la persistencia de datos con la que se cuenta.

La aplicación tiene la capacidad de almacenar los datos de configuración y la información de los trabajos enviados a la red Grid en la máquina local para poder disponer de dichos datos entre distintas sesiones si el usuario lo desea.

De esta manera aunque se cierre el plugin y la hoja de cálculo de Microsoft Excel, quedarán almacenados los datos para poder seguir trabajando con ellos cuando el usuario vuelva a arrancar la aplicación.

Por un lado, se almacenarán en un archivo “mappings” los datos de configuración: tanto los de la configuración de la conexión con la red Grid como los de la configuración del último trabajo enviado (nombre del trabajo, ejecutable, y rango de celdas de entrada).

Al guardar los datos de configuración de la conexión con la red Grid, se ofrece la posibilidad al usuario de guardarlos en su totalidad, de guardar todo excepto las contraseñas, o de no guardarlos.

Por otro lado, se almacenarán en un archivo “trabajos” la información referente a los trabajos enviados a la red Grid. Se almacenan tanto los trabajos que se encuentran en estado

pendiente como los que ya han sido completados pero aún no se ha visualizado el resultado. Una vez que se visualizan los resultados de un trabajo en la hoja de cálculo, éste se borra del fichero.

La información que se guarda de cada uno de los trabajos será su ID en la red Grid, su nombre, el número de tareas totales, el número de tareas completadas hasta el momento y la fecha de lanzamiento.

De esta manera cada vez que se arranca la aplicación, se cargarán dichos ficheros de datos rellenando los campos oportunos en el formulario. Del mismo modo, el fichero de trabajos estará disponible para cuando el usuario quiera realizar una monitorización de los trabajos.

Una vez conseguidas todas estas funcionalidades, se ha conseguido un completo sistema de ejecución de funciones complejas, con alta demanda de capacidad de ejecución, a través de un acceso a red grid transparente para el usuario de la aplicación Microsoft Excel.

2 LAS REDES GRID

2.1. Introducción

¿Qué es Grid?

Llamamos grid al sistema de computación distribuido que permite compartir recursos no centrados geográficamente para resolver problemas de gran escala, de forma uniforme, transparente, segura, eficiente y fiable. Los recursos compartidos pueden ser Ordenadores (PCs, estaciones de trabajo, supercomputadoras, PDA, portátiles, móviles, etc.), software, datos e información, instrumentos especiales (radio, telescopios, etc.), personas/colaboradores...

Podemos ver el grid como una tecnología análoga a las redes de suministro eléctrico: ofrece un único punto de acceso a un conjunto de recursos distribuidos geográficamente en diferentes dominios de administración.

Motivación

Para entender la existencia de este tipo de tecnología, vamos a ver primero el contexto por el cual fue necesaria su creación.

Con el paso del tiempo, el desarrollo del hardware llevó a la mejora y abaratamiento de este, dando paso a tecnologías que antes era impensable emplear, debido a costes o el bajo rendimiento que ofrecían, como lo eran los clusters. Los clusters surgieron como una nueva alternativa muy recomendable en el campo de la computación, tanto paralela como distribuida. Consiste en unir un conjunto de computadores personales o estaciones mediante una red (LAN, SAN, ...). Su principal ventaja es la mejor relación de coste a rendimiento, sin embargo la gran demanda tanto de computación como de espacio y gestión de almacenamiento requeridos por un gran número de aplicaciones que gestionan grandes cantidades de datos y han de hacerlo de forma eficiente y sencilla, exige el uso de nuevas tecnologías, como es el caso de la computación grid o grid computing.

La idea de grid está enfocada fundamentalmente al acceso remoto a recursos computacionales y pretende ser un paradigma de desarrollo sin centrarse en una tecnología concreta. La evolución de grid computing se refleja en el avance de la estandarización de esta tecnología (el estándar de Globus Project es el estándar de facto) donde se encuentra definida la arquitectura del grid, los niveles de acceso, los requisitos, los servicios, etc. Esta definición no incluye el uso de una tecnología concreta si no que explícitamente determina como necesario el uso de múltiples tecnologías para el completo desarrollo del grid global.

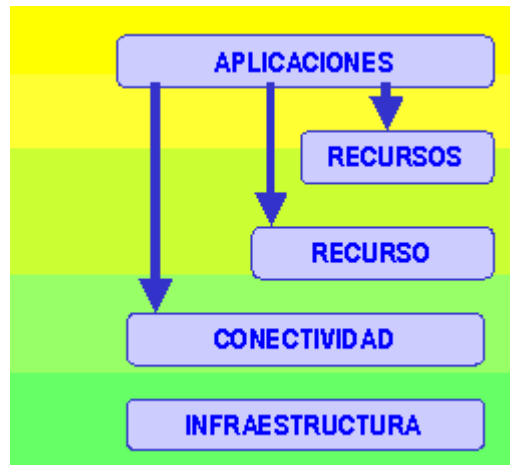
El concepto de grid computing se enmarca dentro de la tecnología de computación distribuida englobando conceptos como sistemas operativos distribuidos, programación multiprocesador, redes de ordenadores, etc. De alguna manera el concepto de grid computing da una unidad conceptual a todos estos problemas de manera que todos ellos pueden verse desde una perspectiva grid.

Además la computación Grid ofrece muchas ventajas frente a otras tecnologías alternativas. La potencia que ofrecen multitud de computadores conectados en red usando Grid es prácticamente ilimitada, además de que ofrece una perfecta integración de sistemas y dispositivos heterogéneos, por lo que las conexiones entre diferentes máquinas no generarán ningún problema. Se trata de una solución altamente escalable, potente y flexible, ya que evitarán problemas de falta de recursos (cuellos de botella) y nunca queda obsoleta, debido a la posibilidad de modificar el número y características de sus componentes.

Estos recursos se distribuyen en la red de forma transparente pero guardando unas pautas de seguridad y políticas de gestión de carácter tanto técnico como económico. Así pues, su objetivo será el de compartir una serie de recursos en la red de manera uniforme, segura, transparente, eficiente y fiable, ofreciendo un único punto de acceso a un conjunto de recursos distribuidos geográficamente en diferentes dominios de administración. Esto nos puede llevar a pensar que la computación Grid permite la creación de empresas virtuales.

Arquitectura

La arquitectura general se articula en cinco niveles: la infraestructura, la conectividad, la gestión del recurso, la gestión de varios recursos y el nivel de aplicación, según se muestra en el siguiente gráfico.



Avance respecto a la World Wide Web

El World Wide Web proporciona un acceso transparente a información que está almacenada en millones de ordenadores repartidos por todo el mundo. Frente a ello, el grid es una infraestructura nueva que proporciona acceso transparente a potencia de cálculo y capacidad de almacenamiento distribuida por una organización o por todo el mundo.

- Proporciona un mecanismo de colaboración transparente entre grupos dispersos, tanto científicos como comerciales.
- Posibilita el funcionamiento de aplicaciones a gran escala.
- Facilita el acceso a recursos distribuidos desde nuestros PCs.
- Todos estos objetivos y beneficios se engloban en la idea de "e-Ciencia".

Aplicaciones

Los beneficios que el grid trae consigo tendrán repercusión en muchos campos:

- Medicina (imágenes, diagnosis y tratamiento).
- Bioinformática (estudios en genómica y proteómica).
- Nanotecnología (diseño de nuevos materiales a escala molecular).

- Ingeniería (diseño, simulación, análisis de fallos y acceso remoto a instrumentos de control).
- Recursos naturales y medio ambiente (previsión meteorológica, observación del planeta, modelos y predicción de sistemas complejos).

Cada uno de los posibles campos de aplicación de la tecnología grid:



Existen cinco grandes áreas de trabajo determinadas por las necesidades de cálculo, espacio para el almacenamiento de los datos y tiempo de respuesta. Las áreas son:

- Supercomputación distribuida. Dentro de esta área se encuentran aquellas aplicaciones cuyas necesidades es imposible satisfacer en un único nodo. Estas necesidades se producen en instantes de tiempo determinados y consumen muchos recursos, por lo que se dice que son puntuales e intensivas. Ejemplo de este tipo de aplicaciones son las simulaciones, las herramientas de cálculo numérico, los procesos de análisis de datos, la extracción de conocimiento de almacenes de datos, etc.
- Sistemas distribuidos en tiempo real. En este tipo de aplicaciones se consideran aquellas que generan un flujo de datos a alta velocidad que debe ser analizado y procesado en tiempo real. Ejemplo de este tipo de aplicaciones son los experimentos de física de alta energía, control remoto de equipos médicos de alta precisión y precio, todos los procesos de la denominada e-Medicine, el tratamiento de imágenes para la visión artificial, etc.
- Proceso intensivo de datos. En esta área nos centramos en aquellas aplicaciones que hacen un uso intensivo del espacio de almacenamiento. Las necesidades de

almacenamiento de este tipo de aplicaciones desbordan la capacidad de almacenamiento de un único nodo y los datos son distribuidos por todo el grid. Además de los beneficios por el incremento de espacio, la distribución de los datos a lo largo del grid permite el acceso a los mismos de forma distribuida. Ejemplos de este tipo de aplicaciones son todos los sistemas gestores de bases de datos distribuidas.

- **Servicios puntuales.** En esta área, nos olvidamos del concepto de potencia de cálculo y capacidad de almacenamiento para centrarnos en recursos que una organización puede considerar como no necesarios. De esta manera el grid ofrece a la organización esos recursos sin que la organización deba desarrollarlos por si misma. Ejemplos de este tipo de aplicaciones son aquellas que permiten acceder a hardware muy específico (equipos costosos de medida o de análisis de muestras) para la realización de labores a distancia.
- **Entornos virtuales de colaboración.** Esta área está relacionada directamente con el concepto de Teleinmersión, de manera que se utilizan los enormes recursos computacionales del grid y su naturaleza distribuida para generar entornos virtuales 3D distribuidos.

2.2. MiddleWare (Globus)

Cuando hablamos de Globus estamos hablando de un conjunto de librerías, servicios y API's para construir Grids. Se trata de un software de tipo Middleware, no de una herramienta de usuario ni una aplicación. Se trata de una tecnología de carácter básico para el Grid puesto que es aquella que permite compartir determinados recursos entre usuarios a través de la red sin perder un ápice de autonomía local.

Con el paso del tiempo ha ido convirtiéndose casi en un estándar para la computación en redes Grid. De hecho es la columna vertebral de gran cantidad de proyectos de ciencia e ingeniería acerca de Grid en muchas universidades alrededor de todo el mundo y parte importante en nuevos productos que comercializan grandes empresas de investigación destinados a llegar al gran público.

A la hora de desarrollar nuevas aplicaciones el conjunto de herramientas ofrecido por Globus puede ser usado o bien de manera independiente o conjuntamente con otros entornos de desarrollo con el fin de potenciar las creaciones obtenidas con estos últimos. Globus contiene numerosos servicios y librerías que resultan necesarios para diversas tareas a realizar con los recursos, como por ejemplo su búsqueda, su gestión e incluso su monitorización. Estos componentes se utilizan también para la gestión de aspectos relacionados con la seguridad y

el tratamiento de ficheros (Grid Resource Management Protocol).

Seguridad.

Archivos (Acceso Global al Almacenamiento Secundario, Global Access to Secondary Storage y FTP en Grid, GridFTP).

Cada organización tiene sus propios modos de trabajo y la colaboración entre organizaciones se ve en muchos casos limitada por las incompatibilidades que puedan existir entre los distintos recursos (tanto a nivel de redes como de equipos individuales o incluso de archivos). Por lo tanto a la hora de plantearse la gestión de multitud de recursos pueden surgir ciertas dudas acerca de las incompatibilidades que puedan presentar estos; Globus elimina dichas obstrucciones y proporciona al usuario que utiliza los servicios ofrecidos una perspectiva sobre el ámbito de trabajo remoto similar a la que obtendría realizando un acceso local a su equipo, manteniendo siempre un cierto nivel de privilegio sobre nuestros recursos. Esto se consigue gracias a los protocolos, interfaces y servicios del núcleo que se ofrecen.

Globus es un código puramente Open-source, lo cual permite que esté actualizado de manera continuada y sea capaz de utilizar las tecnologías más novedosas e innovadoras que aparecen gracias a la colaboración de toda la comunidad que trabaja.

¿Qué cubre Globus Toolkit?

A continuación se explican las capas que cubre Globus Toolkit.



“Connectivity Layer” o capa de conectividad, que se encarga de:

Gestión de servicios (WSRF)

Monitorización/Descubrimiento (WSRF and MDS)

Seguridad (GSI and WS-Security)

Comunicación (XIO)

“Resource Layer” o capa de recursos, que se encarga de:

Gestión de ejecución (GRAM)

Gestión de datos (GridFTP, OGSA-DAI)

En desarrollo: Telecontrol (NTCP/GTCP)

“Collective Layer” o capa colectiva, que se encarga de:

Gestión de datos (RLS, MCS, OGSA-DAI)

Monitorización/Descubrimiento (MDS)

Seguridad (CAS)

Entorno de desarrollo

Desarrollo de Web services

Desarrollo de aplicaciones

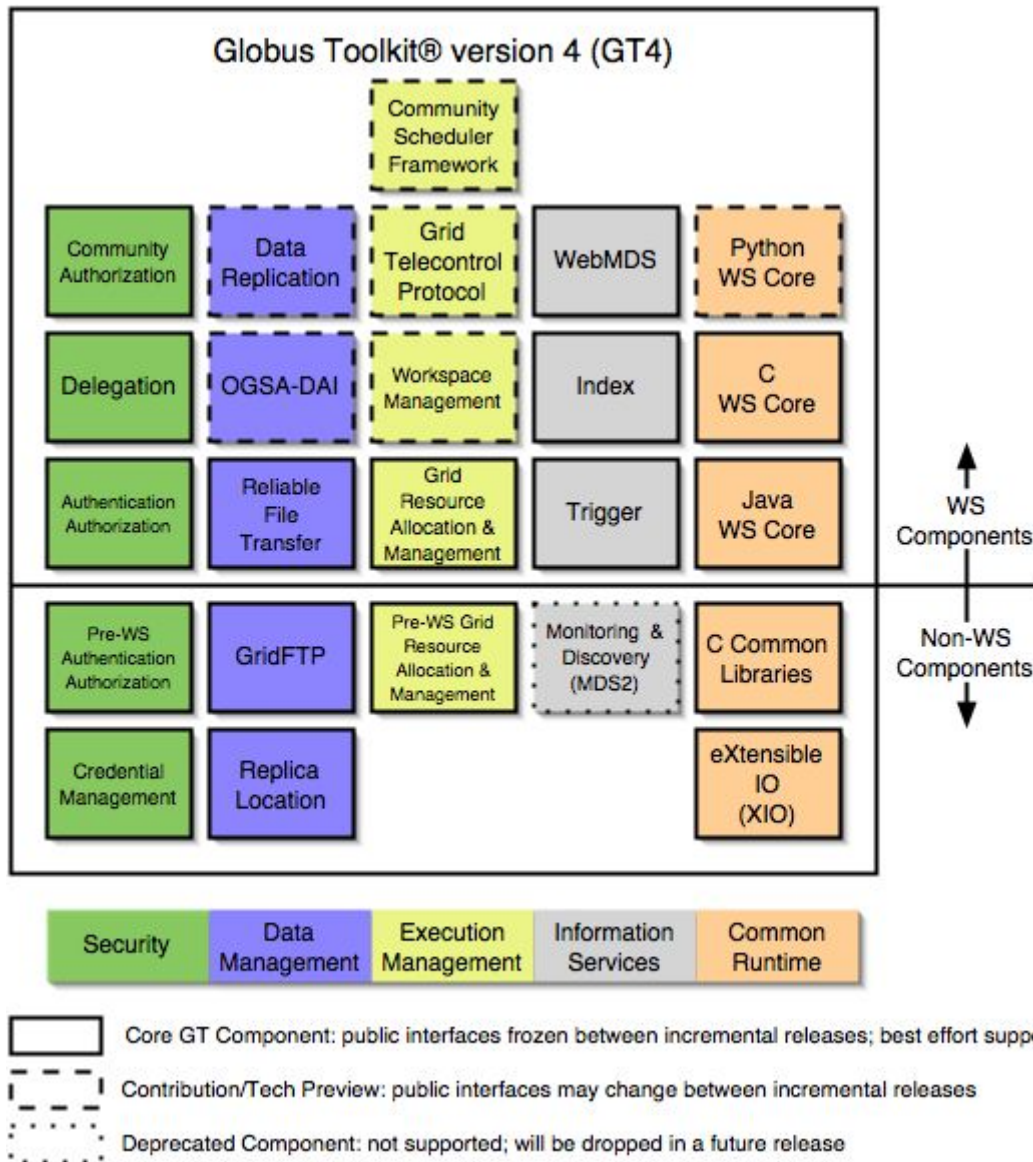
Aplicaciones/servicios seguros

Servicios básicos

Los servicios básicos que ofrece Globus son gestión de ejecución, transferencia de datos, acceso a base de datos, gestión de datos y monitorización y descubrimiento.

Componentes GT4

Sus principales componentes se muestran en la siguiente figura:



- La mayoría de los componentes están implementados con servicios web (+ WSRF).

Los Web Services son aplicaciones basadas en Internet que interactúan de forma dinámica con otras aplicaciones basadas en Web utilizando estándares como XML, UDDI y SOAP. Tanto IBM, como Microsoft y Sun están decididas a unir la tecnología grid con los Web Services. Mediante la última versión de Globus Toolkit se pueden encontrar grid services en la red. También posibilita la creación de una interfaz para el servicio y permite invocar una instancia de un grid service.

- Unos cuantos componentes no utilizan servicios web.

Componentes que, por su naturaleza, no tiene sentido implementarlos con servicios web. P.ej. herramientas para gestión de credenciales. Componentes que se mantienen por

compatibilidad hacia atrás. Algunos de estos componentes han sido reimplementados con servicios web, y serán eliminados en futuras versiones

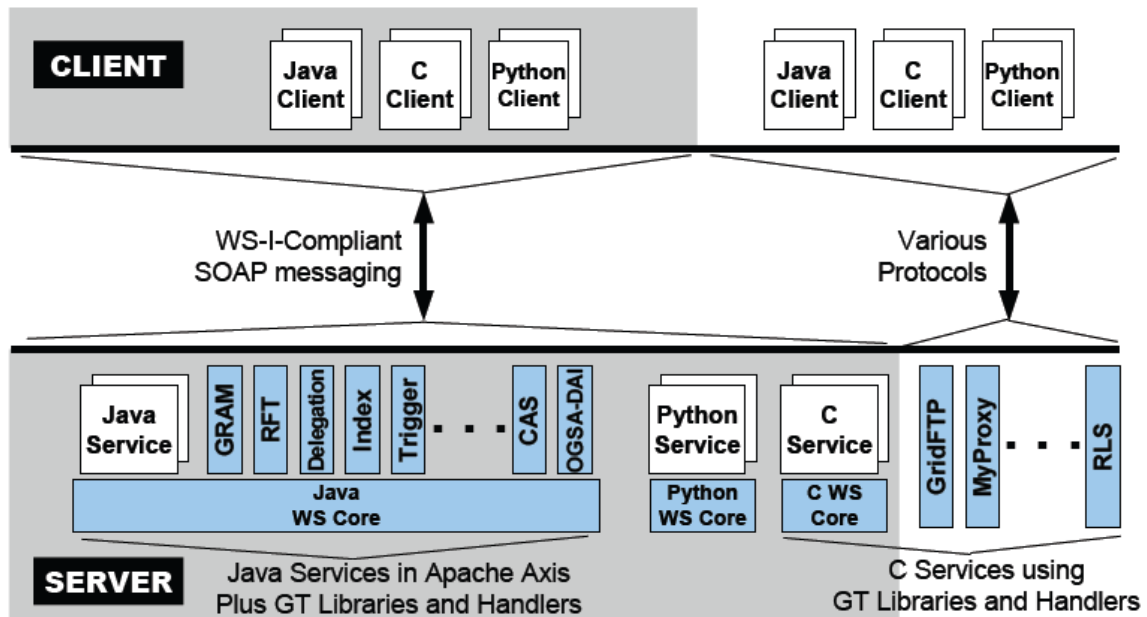


Figura: Entornos Grid Basados en el Globus Toolkit 4

Utilización del middleware Globus Toolkit

Definitivamente casi no tiene ningún valor para los usuarios finales.

- Dispone de Muy poco interfaz de usuario (final)
- No es una herramienta que se instala y está lista para explotación por parte de científicos/ingenieros/clientes

El Globus Toolkit es útil para desarrolladores de aplicaciones e integradores de software.

- Hay que tener clara la aplicación que queremos desarrollar.
- Hay que tener la experiencia adecuada.
- Hay que tener el hardware/software necesario.
- Requiere planificación.

La computación en grid es actualmente un fenómeno en emergencia al que Globus Project (el proyecto iniciador) ha proporcionado una implementación estándar que es utilizada por multitud de investigadores y empresas.

2.3 Metaplanificación (GridWay)

Grids ofrece un dramático crecimiento en el número de recursos de procesamiento y almacenamiento disponibles que pueden ser distribuidos a las aplicaciones. Sin embargo, el envío eficiente de trabajos y su gestión continúan siendo difícilmente accesibles a los científicos e ingenieros comunes debido a su naturaleza dinámica y compleja. La meta del proyecto Gridway es investigar y desarrollar la tecnología requerida para realizar automáticamente todos los pasos de envío de trabajos y también proporcionar los mecanismos de tiempo de ejecución requeridos para adaptar dinámicamente la ejecución de la aplicación.

El marco de trabajo Gridway ha sido desarrollado para reducir el vacío entre el middleware Grid y los desarrolladores de aplicaciones.

El marco de trabajo Gridway es un componente para la metaplanificación en el ecosistema Grid, previsto para usuarios finales y desarrolladores de aplicaciones Grid. Es un gestor de carga que realiza la gestión de la ejecución de trabajos y la repartición de recursos en un Grid formado por distintas plataformas de computación dirigidas por servicios Globus. Gridway permite desatendidos, confiables y ejecución eficiente de trabajos simples, complejos o en array en grids heterogéneos y dinámicos. Gridway realiza toda la planificación y envíos de trabajos de forma transparente al usuario final y adapta la ejecución de trabajos para cambiar las condiciones grid proporcionando mecanismos de recuperación de fallos, planificación dinámica. Gridway en Globus permite desacoplar las aplicaciones y los sistemas de gestión local subyacentes.

Gridway es un metaplanificador que permite una compartición de recursos computacionales (clusters, servidores, supercomputadoras...) de manera segura, eficiente y a larga escala, gestionado por diferentes sistemas LRM (Local Resource Management), como PBS, SGE, LSF, Condor..., con una organización simple o dispersada a través de varios dominios administrativos.

Gridway es un proyecto Globus, que cuenta con la filosofía Globus y con las directrices para el desarrollo colaborativo y contribuciones tanto de individuales como de corporaciones a lo largo del mundo.

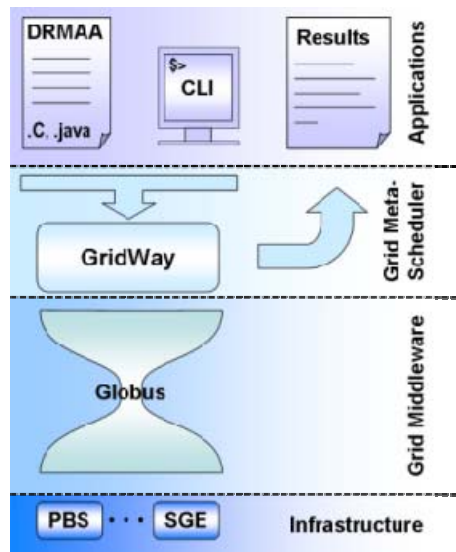
Existe un número de sistemas de gestión y planificación de la carga de trabajo tanto comerciales como de libre distribución disponibles hoy en día, cada uno adecuado para diferentes infraestructuras de computadores subyacentes y perfiles de ejecución. Gridway destaca entre otros sistemas metaplanificadores porque ha sido diseñado específicamente para trabajar por encima de los servicios de Globus, ofreciendo la más alta funcionalidad, calidad de servicio y confiabilidad en esta clase de infraestructuras.

Cuenta con muchas ventajas para:

- **Directores de infraestructuras y proyectos:** Gridway es un proyecto comunitario de libre distribución, que cuenta con la filosofía Globus y las directrices para el desarrollo colaborativo y no necesita de nuevos servicios, sino que trabaja con los mismos que Globus.
- **Integradores de sistemas:** Gridway es altamente modular, permitiendo la adaptación a diferentes infraestructuras grid y soporta varios estándares OGF. También es fácilmente extensible y podría ser utilizado o extendido por arquitecturas más complejas que implementen los acuerdos a nivel de servicio.
- **Jefes de sistemas:** Gridway proporciona un marco de trabajo de planificación similar al encontrado en los sistemas de gestión de recursos locales (LRM), permitiendo llevar la cuenta de los recursos y definiendo políticas de planificación.
- **Desarrolladores de aplicaciones:** Gridway implementa el API estándar DRMAA (C y Java), asegurando la compatibilidad de aplicaciones con sistemas LRM que implementan el estándar como SGE, Condor, Torque, ...
- **Usuarios finales:** Gridway permite a los usuarios finales enviar, monitorizar, sincronizar y controlar trabajos de forma totalmente transparente y similar los sistemas de gestión de recursos locales, que podrían ser descritos usando el estándar OGF JSDL. El planificador les facilita el trabajo ya que puede gestionar diferentes situaciones de fallo de forma automática.

Con Gridway, una infraestructura Grid puede ser explotada y gestionada de la misma manera que un cluster local.

Su arquitectura se detalla en la siguiente figura:



Se requiere una gestión de la carga de trabajo en el lado del cliente para proporcionar al usuario final unos paradigmas de programación portables e interfaces comunes. Gridway es una herramienta de gestión de ejecución para el uso de Globus Toolkit. Es un middleware que usa Globus, y que puede ser usado en cualquier infraestructura Grid basada en Globus como núcleo de grid.

3 PANORAMA ACTUAL

3.1 Necesidades actuales

Posibilidades de Grid en el entorno actual

En la actualidad no basta con utilizar un ordenador para dar solución a las necesidades diarias, se busca una progresiva mejora para potenciar los resultados obtenidos. Las máquinas son cada día más potentes pero esto en ocasiones no es suficiente para los objetivos que buscamos; bien por limitaciones de memoria, bien por no llegar a una capacidad de proceso mínima, en ocasiones hay procesos que resultan inviables a la hora de ser planteados en las máquinas tradicionales que conocemos, llegando a ser imposible su desarrollo incluso en las máquinas más potentes. Además, en muchas ocasiones los recursos de los que se disponen se encuentran infrautilizados. Según estudios realizados los servidores de la gran mayoría de las empresas solamente utilizan el 60 por ciento de su capacidad, dejando un amplio margen que está en desuso y puede resultar muy útil para realizar ciertos procesamientos de forma paralela a su explotación normal.

Es en los casos mencionados anteriormente dónde encuentran su utilidad las redes GRID. Mediante la partición y distribución de tareas entre distintos equipos se obtienen una mejora en el procesamiento que se hace difícil de alcanzar a través de una única máquina. A esto hemos de sumarle la rapidez que aportan este tipo de servicios.

Otra ventaja que aporta el GRID, sobre todo en el ámbito empresarial, radica en el aspecto económico, puesto que ofrece la posibilidad de contar con una mayor potencia de procesamiento sin necesidad de gastar grandes cantidades en equipos que estén a la altura del rendimiento esperado. La implantación de un sistema basado en el GRID puede generar un ahorro de entre el 8,5 y el 20 por ciento en los presupuestos de tecnologías de la información de las compañías.

Hoy en día los principales impulsores de la tecnología GRID son las universidades y los centros de investigación y desarrollo, los cuales intentan aprovechar los avances conseguidos para su aplicación en muy diversas áreas, como por ejemplo la medicina y la salud o ciertos estudios enfocados al ámbito de determinados campos de la ciencia (física de altas energías, organización virtual de química computacional,...). Pero los avances en el GRID no se quedan ahí y determinados sectores empresariales ya han empezado a ponerlos en práctica para obtener un mejor rendimiento en sus negocios; grandes empresas e instituciones como Telefónica, El corte ingles, GMV o la agencia espacial europea e incluso entidades bancarias

(Banesto, BBVA, Santander Central Hispano) aplican tecnologías GRID en algunas de sus áreas de trabajo.

Las hojas de cálculo como soporte de trabajo

Las hojas de cálculo son un instrumento de trabajo multifuncional que gozan de gran aceptación dentro de diversos campos, convirtiéndose en potentes herramientas de trabajo que se han convertido en muchos casos en necesarias e imprescindibles. Su sencillez de uso y su amplitud de recursos hacen que sean usadas para muy diversos fines, desde llevar un pequeño control de la contabilidad a nivel personal hasta grandes proyectos a nivel empresarial.

Las hojas de cálculo aparecieron en 1979 de la mano de VisiCalc, una aplicación creada por Dan Bricklin para los computadores Apple II. En aquella época eran consideradas como aplicaciones de alto nivel (“cuarta generación”) y suponían un claro avance permitiendo olvidar algunos métodos tradicionales de contabilidad en pos de un instrumento que conseguía una automatización de las tareas más pesadas y tediosas. Además supuso una herramienta que abría el camino a una planificación e inversión de los trabajos más pesados, lo que antes era muy difícil de imaginar. Estos programas puestos en manos de algunos expertos permitieron que las empresas mejorasen sus rendimientos y que, gracias a la colaboración entre estos expertos y los desarrolladores, las hojas de cálculo experimentasen una importante y progresiva mejora de sus funciones y de su rendimiento.

Las hojas de Cálculo proveen un entorno de trabajo orientado a diversas facetas puesto que permiten desarrollar problemas de muy distinta índole como por ejemplo modelado, matemáticas, ingeniería o ciencias sociales. Esto es debido en parte a que no se centran en lo que puede conseguir la herramienta sino en como pueden ayudar al usuario a obtener mejores soluciones a sus problemas.

Hoy en día podemos encontrar diversas programas de gestión de hojas de cálculo en el mercado: OpenOffice, Lotus 1-2-3, Google spreadsheets (hojas on-line), ... Aunque la aplicación de hojas de cálculo más comúnmente utilizada es Excel. En 1985 Microsoft lanzó su particular visión de estos programas y en la actualidad cuenta con el 90 % aproximado de cuota de mercado.

Una de las funciones que aporta MS Excel es la de trabajar como una base de datos, permitiendo un manejo fluido de grandes cantidades información. Esto resulta de gran utilidad debido a que, combinando la variedad de las funciones disponibles y esta capacidad de almacenamiento, obtenemos una herramienta potente y versátil; pero bien es cierto que las necesidades a la hora de gestionar dicha información aumentan con el paso del tiempo e

intentar trabajar con muchos datos en un mismo instante puede acabar saturando nuestras máquinas.

En cada nueva versión de MS Excel que aparece en el mercado su capacidad de cálculo y de gestión de información aumenta. En la última versión disponible, la duodécima (más conocida como Excel 2007), la hoja de cálculo se ha ampliado de 224 a 234 celdas, o para ser más exactos, de 256 columnas x 65.536 filas (28 x 216) a 16.384 x 1.048.576 (214 x 220). Este incremento del tamaño de cada hoja hace pensar que será necesaria alguna ayuda adicional a los mecanismos de proceso del propio Excel a la hora de manejar grandes cantidades de información en operaciones complejas. Es por esto que se desarrollan diversidad de complementos para MS Excel que le ayudan a aligerar las cargas de trabajo de muy diversas maneras: desde la aplicación de complejos algoritmos a pesadas tareas para aligerar los procesos hasta la distribución de las tareas entre distintas entidades (como por ejemplo puede ser nuestro caso).

Microsoft provee a los programadores que quieran trabajar con Excel de un modelo de objetos que puede ser usado para controlarlos de una forma programada. Este modelo consta de un conjunto de clases y métodos que son los homólogos de sus correspondientes representaciones lógicas en Excel: libros, hojas, celdas, ... Este proceso de control permite que se puedan crear multitud de accesorios que posean un determinado grado de control sobre la información que se maneja, de manera que se puedan manipular elementos propios de Excel para generar nuevos libros, hojas, tablas, o simplemente tomar o introducir datos de estos.

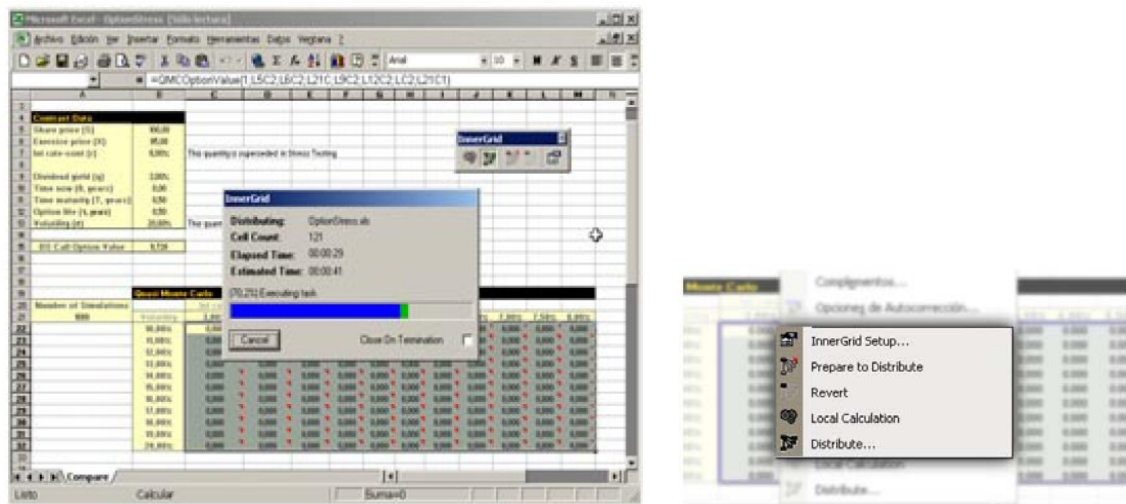
3.2 Soluciones existentes

Actualmente hay varias extensiones de Excel disponibles que interactúan con las redes GRID. Algunas de ellas son las siguientes:

Excel Connector

Esta es una ampliación de las funcionalidades de Excel que trabaja directamente con InnerGrid, una herramienta enfocada a la utilización del GRID por parte del mundo empresarial. Su planteamiento es el de ayudar a resolver los cálculos que contienen complejas fórmulas de manera más rápida y eficaz, trabajando de forma completamente transparente al usuario final y sin la necesidad de cambiar el método de trabajo que se aplicaba antes de utilizar esta extensión. Para su utilización el usuario dispone de nuevas opciones en algunos menús contextuales de MS Excel así como de una barra de herramientas específica para la resolución de los cálculos en el GRID.

Esta herramienta es utilizada actualmente por diversos organismos como por ejemplo numerosas entidades bancarias (Santander Central Hispano,...).



Aspecto que presenta Excel Connector al usuario

Platform Symphony adapter for MS Excel

Este software está enfocado hacia un mercado económico, para una toma de decisiones más rápida y efectiva a la hora de la concesión de créditos o el estudio de riesgos de mercado, reduciendo de manera considerable el tiempo de proceso a la hora de mostrar resultados al usuario. Proporciona una mejora lineal de las simulaciones y garantiza una entrega de los resultados solicitados en el tiempo previsto debido a un sistema de distribución del tiempo entre los cálculos más complejos. Esto es así gracias a la combinación de la ejecución local y remota de las tareas, de tal manera que se asegura un sistema de turnos rotativo para todos los procesos GRID activos.

Un aspecto clave es la gestión de los XLLs. Un XLL es un fichero que define la forma en la que los documentos XML nativos de Excel deben conectarse entre si, un modo estandar par representar enlaces entre recursos. Con este lenguaje se pueden implementar enlaces extendidos o entre múltiples recursos, aparte de los enlaces simples que provee HTML. Al proveer de un único punto de acceso a los ficheros XLL se protege sobremanera aquellos aspectos de la propiedad intelectual que resultan más críticos. Además, la gestión centralizada y bajo un estricto control permite una actualización continua de los XLLs que asegura la disposición de su versión más actual. Estos se almacenan y ejecutan en la base de procesamiento de Platform Symphony, permitiendo a los administradores trabajar con las distintas versiones que existan y reduciendo tanto la complejidad como los riesgos de su tratamiento.

Al igual que el resto de aplicaciones presentadas, los archivos de Excel con los que se esté trabajando no necesitan de ningún tipo de reprogramación ni la aplicación de ningún tipo de cambio para que puedan utilizar el adaptador. El código de los XLLs es traducido directamente y enviado como si se tratase de la propia base de trabajo de Platform Symphony sin necesidad de reescribir el plug-in en ningún lenguaje de programación.

ActiveSheets

Este plug-in se sustenta en otras dos aplicaciones desarrolladas por el mismo equipo denominadas Nimrod y EnFunzion. Cuando un usuario plantea un trabajo a estas aplicaciones, ellas generan un fichero declarativo con el plan a seguir, el cual recoge los parámetros, los valores por defecto y los comandos necesarios para realizar el trabajo. La máquina desde la que se invoca el trabajo se convierte en la máquina “raíz” siendo esta la que controla el experimento. El sistema, con la información de la que dispone, lanza el trabajo a la primera máquina disponible; cada nodo de trabajo puede disponer de un tipo de tecnología distinta. Las dos aplicaciones de apoyo dejan en manos del usuario el acceso y posterior tratamiento de los resultados; Activesheet es la herramienta que automatiza este proceso mediante la combinación de todas las funciones de Nimrod dentro de una herramienta para la hoja de cálculo.

Las funciones de tratamiento de la información están implementadas con Visual Basic for Applications (VBA) como stubs (un pequeño ejecutable que acapara parte del control al crear un trabajo), mediante la aplicación de automatización OLE. Estos stubs aportan argumentos e identificadores de funciones a un DLL de ActiveX el cual se comunica con la máquina remota para presentarle la información pertinente al usuario. Una vez obtenido el resultado este será devuelto mediante este control. La clave de este trabajo está en una tabla que almacene el estado de cada trabajo solicitado y que sirve, en cierta medida, como una caché.

Esta aplicación presenta actualmente la posibilidad de que el usuario decida que funciones deben de ser ejecutadas en paralelo en la máquina remota. La implementación da lugar a dos tipos distintos de paralelismo: a nivel de conjunto de celdas e internamente dentro de una misma celda.

El sistema coexiste con MS Excel sin que exista la necesidad de realizar ninguna modificación en su uso ni de reescribir las hojas de cálculo y trabaja con las versiones 97 y 2000 de la aplicación.

ExcelGrid

ExcelGrid es el plug-in actual más completo y potente para Excel que utiliza los recursos de las redes GRID. Se trata, en concreto, de un sistema middleware a nivel de usuario. Colabora

con el programa extendiéndolo para participar en redes corporativas o globales obteniendo los consiguientes beneficios de un procesamiento más veloz. Provee de un medio de conexión al GRID desde una hoja cualquiera de Excel sin la necesidad de realizar ninguna modificación sobre esta; además permite la conexión tanto a redes globales construidas utilizando conjuntamente Gridbus con Globus como a redes propias creadas usando Alchemi mediante un proceso de configuración según la elección del usuario.

Con ExcelGrid pueden ejecutarse aplicaciones externas a Excel que toman como parámetros los valores contenidos en las celdas seleccionadas de la hoja en la que se esté trabajando. Estas aplicaciones son enviadas a equipos remotos en los cuales serán ejecutadas. El lanzamiento de nuevas tareas se realiza a través de un sencillo GUI que permite realizar diversas funciones con estas, desde su creación hasta su eliminación pasando por la visualización de su estado o la recuperación de los resultados obtenidos para su volcado sobre la hoja en la que estemos trabajando.

Para llegar a conseguir una conexión ExcelGrid utiliza Alchemi. Este es un framework destinado al trabajo de redes GRID basado en Windows e implementado con tecnología Microsoft .NET que provee una maquinaria de ejecución y un entorno de programación destinados a la creación de redes GRID personales y el desarrollo de aplicaciones destinadas a estas redes. Soporta la programación de aplicaciones para GRID orientadas a objetos así como la ejecución de aplicaciones multiplataforma, siempre a través de servicios web. Los aspectos claves que aporta Alchemi son la creación de clusters de ordenadores personales que no poseen un espacio de memoria compartida gracias a internet, la utilización de clusters para crear GRIDs cooperativos y con una determinada jerarquía implícita, ejecución de tareas de forma dedicada o no dedicada mediante clusters o nodos individuales, un modelo de programación orientada a objetos para la creación de hilos dentro de GRID (mediante una programación de grano fino) destinado a una interacción entre distintas plataformas.

Por otra parte, la otra herramienta utilizada por ExcelGrid es Gridbus. Este es un framework de trabajo multiplataforma que proporciona ciertos servicios que permiten la ejecución de tareas en diversos sistemas de middleware que trabajan a bajo nivel como por ejemplo Globus o UNICORE. Para esta aplicación Gridbus trabaja como un mero servidor de aplicaciones, manteniéndose atento a las conexiones y ofreciendo servicios de 'brokering' en las redes GRID globales. Los mensajes que envía este servidor son captados por el Cliente Grid el cual los utiliza para actualizar el estado de los trabajos planificados.

ExcelGrid está formado por los siguientes componentes principales:

- Add-In: Es el encargado de la creación del acceso al plug-in desde Excel, el frontal de comunicación con el modelo de aplicación de Excel, creando un acceso en el menú de Herramientas para que el usuario pueda acceder. Se trata de una aplicación de

extensibilidad que implementa el interfaz `Extensibility.IDTExtensibility2`, el cual permite que el acceso al plug-in esté presente cada vez que se abre MS Excel.

- **Runner:** este es el GUI con el que interactúa el usuario, el interfaz principal desde el que va a poder configurar las opciones solicitadas para la preparación de nuevos trabajos, así como el envío de estos. Además nos informa del estado en el que se encuentra cualquier trabajo en ejecución. Se trata de una aplicación estándar de Windows creada con una plantilla de proyecto creada con Visual Studio .NET. Contiene una clase que da forma al GUI con el que interactúa el usuario, otra que permite el almacenamiento de los detalles de configuración del usuario de forma que esta permanezca a través de distintas y sucesivas ejecuciones de ExcelGrid, y la clase principal que inicia la ejecución de este componente.

El cliente GRID intenta conectarse con el middleware enviándole datos tomados directamente de la hoja de Excel, transformados en unidades de información adecuadas para su utilización en el GRID. La comunicación se realiza mediante una conexión directa entre sockets y un protocolo específico para la especificación de los trabajos y sus parámetros. Estos últimos son enviados como un único mensaje que el servidor se encarga de seccionar para crear los trabajos. De esta manera se crea un paralelismo a nivel de trabajos.

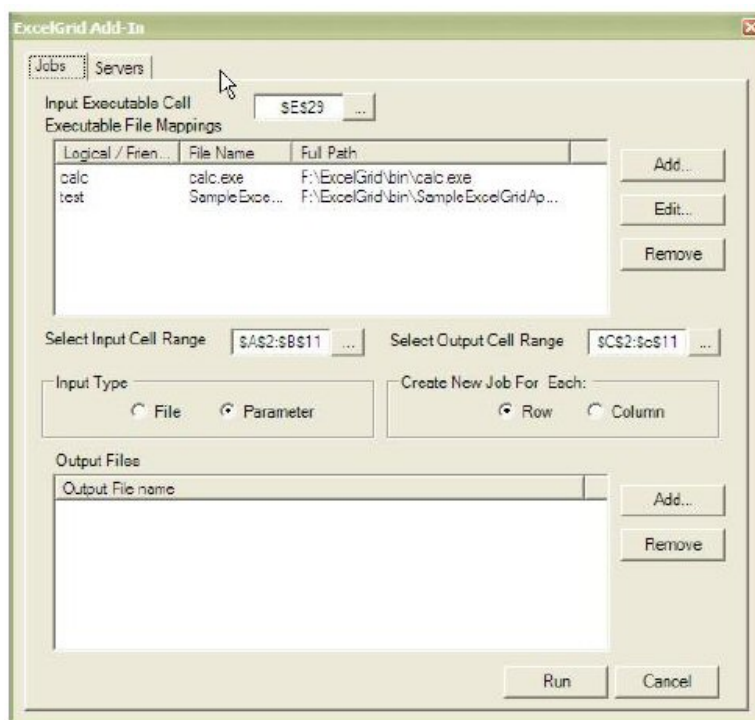
Los distintos componentes de ExcelGrid han sido desarrollados pensando en un diseño orientado a objetos, implementándolos con lenguajes muy diversos como Visual Basic o C# (.NET) y Java.

Tanto en el caso de trabajar con Alchemi como con Gridbus el proceso a la hora de comenzar a trabajar es el mismo: el usuario solicita la ejecución de un trabajo y el Cliente Grid prepara este para enviarlo. Cuando se tratan de redes locales envía el trabajo, Alchemi lo recibe a través de su API y pone en marcha unas funciones de aviso de tal manera que, cuando alguno de los trabajos recibidos falla o finaliza se pone este hecho en conocimiento del cliente, el cual actualiza la información que posee.

En el caso de tener que utilizar Gridbus el método es distinto: el broker está implementado en Java y el Cliente Grid en C# por lo que una comunicación directa entre ambos es imposible, hecho por el que se utilizan sockets. El cliente se mantiene conectado al servidor mientras dura la ejecución del trabajo y va recibiendo mensajes de este acerca de si ha acabado o ha sufrido algún fallo. Al captar algunos de estos mensajes se avisa al Runner para que actualice la información que se muestra al usuario.

Como ya dijimos anteriormente Excel posee un modelo de objetos que permite al programador interactuar con los elementos básicos de cualquier documento que se cree. El grado de control existente es posible gracias a la tecnología de Microsoft COM (Component

Object Model). Si bien COM no puede comunicarse de una forma directa con la tecnología .NET, esta última posee una característica llamada COM Interop que proporciona un entorno de llamadas que permite un cierto grado de interconexión entre ambos. De esta manera los componentes de ExcelGrid se comportan como si se tratase de miembros .NET nativos, pudiendo trabajar con el modelo de objetos de Excel, gracias al uso de la citada característica.



Aspecto del interfaz de usuario (GUI) de ExcelGrid

Aquellos cálculos y procesos que necesitan ser realizados en los equipos remotos son representados como trabajos. Cada uno de ellos lleva asociado un ejecutable que es el encargado de realizar el proceso correspondiente en los nodos; además contendrá una serie de información de entrada y de salida (parámetros y resultados). Cada conjunto de información de este tipo que se quiere transmitir se encapsula en un determinado paquete dependiendo del tipo de GRID a utilizar, y enviándolo a la red como un objeto. Este objeto contiene la información de aquellas celdas que van a servir como parámetros de entrada, la cual se transforma previamente al envío para crear un paquete del tipo adecuado (GParameter object).

Previo al envío de los datos el cliente GRID (Alchemi o Gridbus) establece una comunicación con el entorno en el que se va a desarrollar el trabajo. El cliente GRID envía al nodo principal de la red tantos mensajes como trabajos sean lanzados y se mantiene a la escucha, esperando mensajes o eventos provenientes de ese nodo, caso en el que se lanzará una alerta indicando

el evento sucedido que será recogida por una entidad del componente Runner. Esta entidad recoge los resultados obtenidos y los traslada a la hoja de Excel. Todo este proceso se realiza de manera automática gracias a los frameworks que utiliza la aplicación (Alchemi y Gridbus broker).

Previamente a la creación de un trabajo el usuario debe de indicar a través del GUI una serie de datos a la aplicación: rango de celdas con los datos de entrada que servirán de parámetros para el trabajo, localización del ejecutable que será enviado para el procesamiento de los datos de entrada y nombre por el que se identificará el trabajo. Además cabe la posibilidad de definir desde el mismo interfaz si los parámetros de entrada indicados serán agrupados por filas o por columnas a la hora de ser asignados a un nuevo trabajo así como el rango de celdas de salida donde se colocarán los resultados que se obtengan en la ejecución de los trabajos.

La mejora del rendimiento experimentada cuando se utiliza ExcelGrid para realizar algunos trabajos es considerable. Cuando el plug-in trabaja con redes gestionadas por Alchemy el incremento del tiempo de ejecución es lineal con respecto a la complejidad de los cálculos a realizar. A la hora de trabajar con un agente GridBus se puede observar que, a medida que crece el número de nodos a los que enviar trabajos, el tiempo que tardan en completarse dichos trabajos disminuye, aunque no lo hace dentro de un rango totalmente lineal, sino que mejora en un grado ligeramente inferior.

3.3 Novedades y mejoras que aportamos

El desarrollo que se ha realizado ha pretendido desde el comienzo ser una herramienta de apoyo a las necesidades más complejas de computación para Excel, pero sin dejar de ser manejable y accesible para cualquier persona. Utilizando una tecnología completamente transparente al usuario final, sin la necesidad de que este posea grandes conocimientos en el aspecto de las redes GRID y sin que tenga que implicarse en exceso en la gestión del plug-in; algo sencillo y funcional. Aquel que acceda a esta aplicación no tiene la necesidad de saber que es lo que está sucediendo internamente en su mecanismo de trabajo, le basta con solicitar la realización de unos determinados cálculos y procesamientos al plug-in y dejar que este haga su trabajo hasta la obtención de algún resultado.

La mayoría de las aplicaciones que hemos descrito anteriormente son extensiones de herramientas de carácter más general que implican diversas áreas de trabajo dentro de la mejora que aportan las redes GRID. Entre esos campos en los que se aplican está el trabajo diario con MS Excel. La nuestra es una aplicación diseñada exclusivamente para Excel, con lo que garantiza que todo el esfuerzo realizado por el cliente GRID que se utiliza va enfocado únicamente a la mejora y aceleración de los cálculos solicitados por este. En este sentido nuestro proyecto presenta en cierta medida alguna similitud con ExcelGrid, desarrollo del que

ya hablamos antes. Si bien es cierto que incluso en apariencia externa e interfaz son muy similares y que su arquitectura posee un cierto parecido, la manera en que trabaja internamente es completamente distinta.

La aplicación que hemos creado utiliza Gridway como cliente GRID. Gridway es un sistema gestor de carga y meta-planificador diseñado para trabajar exclusivamente sobre servicios provistos por GLOBUS, de tal manera que permite trabajar con una red GRID de la misma manera que si se tratase de un cluster de máquinas (la gestión de varios usuarios se puede realizar desde un puesto de administrador). Gridway trabaja en colaboración con otros meta-planificadores de tal forma que no se genere la figura de un único meta-planificador que controle y gestione (y sature) los tránsitos con el cliente; de la misma manera pueden existir diversos usuarios dentro de cada infraestructura. Cada infraestructura solicita la ejecución de determinados trabajos, que compiten con las solicitudes del resto de infraestructuras existentes por los recursos disponibles. Además facilita una gestión centralizada de diversos recursos dentro de un único dominio administrativo, siempre y cuando el administrador pueda aplicar políticas de uso unificadas. El uso de Gridway repara en una serie de beneficios para el usuario final del producto:

- El planificador gestiona todas las situaciones posibles, así como las de fallo, permitiendo al usuario desentenderse de las labores más complejas.
- Asegura una completa compatibilidad con cualquier tipo de aplicación que se ejecute, no se ajusta a un determinado perfil de aplicación ni de entorno predefinido; mediante la transmisión de ficheros y el acceso a las base de datos se garantiza la ejecución de cualquier tipo de tarea.
- Se asegura la ejecución en los nodos disponibles más potentes dentro de la red GRID sin la necesidad de una implicación excesiva del usuario al configurar la ejecución.

A la hora de comunicarnos con el cliente GRID utilizamos un protocolo SSH a través de una librería especializada que incorporamos a nuestro proyecto (Sharp SSH). Esta librería está desarrollada para aplicarla con un lenguaje C#, lo cual es perfectamente acoplable a nuestro diseño gracias a la adaptación que proporciona trabajar con VBA. La ventaja que aporta SSH a nuestro desarrollo es la de proporcionar una comunicación mucho más segura para la transmisión de datos. Además la comunicación es directa entre el equipo que solicita la ejecución del trabajo y el cliente GRID, sin intermediación de ningún otro software.

El plug-in está formado por una única entidad que trabaja directamente entre MS Excel y el cliente GRID, sin necesidad de terceros programas que gestionen conexiones ni manejen los datos que se envían. El componente que forma la interfaz se encarga al mismo tiempo de almacenar los datos necesarios y de gestionar el envío y recepción de los trabajos, formando

una única entidad. De esta forma se evitan innecesarias comunicaciones entre componentes que podrían dar lugar a errores y se simplifica la creación, envío de los trabajos y la recuperación de los resultados.

En gran medida esta unificación de tareas se debe al entorno de desarrollo que hemos utilizado para construir el plug-in. Al trabajar con VBA hemos podido juntar casi toda la funcionalidad en un único componente. Esto se debe a que la tecnología .NET permite la cooperación de distintos lenguajes bajo un mismo entorno de trabajo.

El GUI con el que interactúa el usuario es, además, muy sencillo de utilizar; la funcionalidad que implementa no necesita apenas de configuraciones. Aquel que quiera utilizar el plug-in le basta con tener una cuenta en Gridway y saber sus datos: a partir de aquí el trabajo en Excel es automático. Una vez se introduzcan esos datos el plug-in está listo para comenzar a enviar trabajos. Además esta pequeña iniciación para poder dar comienzo a nuestro trabajo cuenta con la ventaja de que puede ser guardada en el equipo local, de tal manera que en sucesivas ocasiones que lancemos la aplicación podremos acceder directamente a los trabajos ya creados o generar nuevos sin necesidad de tener que recordar todos nuestros datos. Este almacenamiento de nuestra identificación es cifrado por lo que no supone ningún riesgo contra la privacidad. Y si decidimos cambiar de usuario de una sesión a otra, los datos existentes desaparecen, sin comprometer la identidad de ninguno de los que accedan al plug-in.

Otro punto que aporta cierta flexibilidad a la aplicación es la posibilidad de elegir la localización del ejecutable que va a realizar los cálculos solicitados. Este ejecutable puede estar en el equipo local del usuario y ser enviado junto con los parámetros necesarios para su ejecución o bien estar alojado en los equipos remotos a los que se va a transmitir las entradas. Al solicitarse la creación de un nuevo trabajo se ha de indicar a través del GUI dónde encontrar el ejecutable al que aplicar los parámetros de entrada que se pasan.

Un aspecto distinto en nuestro modelo es la forma en la que se gestionan los resultados que se obtienen de la ejecución de los distintos trabajos. La recuperación de estos resultados no se tramita automáticamente, es el usuario el que decide cuando insertarlos en la hoja de cálculo. Mediante una consulta a través del cliente GRID se puede conocer el estado de cada trabajo solicitado y, una vez que estos hayan finalizado, puede enviarse la orden necesaria para trasladar los resultados desde el equipo remoto en el que se estuviese ejecutando el trabajo. Quizá plantee en cierta medida la necesidad de una mayor participación por parte del usuario, el cual podría preferir una mayor automatización del proceso, pero asegura un mayor control por parte de este. De esta manera se pueden gestionar de la forma deseada los resultados obtenidos, descartándolos y eliminándolos directamente del equipo remoto si finalmente no nos resultan necesarios.

Gracias a esta gestión de la información evitamos contar con un número masivo de comunicaciones, al mismo tiempo que ganamos en seguridad. Al no estar preguntando continuamente acerca del estado de los trabajos evitamos posibles intrusiones en la privacidad de los usuarios.

Podemos deshacernos de un resultado que no queramos una vez haya finalizado la ejecución de un trabajo solicitado o podemos cancelar la realización de dicho trabajo en tiempo de ejecución mediante una sencilla orden al equipo remoto a través del GUI con el que interactúa el usuario. Este novedoso procedimiento permite enviar una orden mediante Gridway directamente a los equipos en los que se esté ejecutando una determinada tarea para que cancele cualquier proceso en el que esté involucrada la tarea, sea el estado que fuere en el que esta se encuentre.

4 EL PROYECTO PLUGIN EXCEL GRID

4.1 Planteamiento del proyecto

El objetivo del proyecto Plugin Excel Grid es crear un plugin para la aplicación Microsoft Excel que permita ejecutar funciones sobre celdas y depositar el resultado en una celda destino. La peculiaridad de la ejecución es que los cálculos no se realizan en la máquina local, sino que se realizan en remoto, liberando así la carga de la ejecución a la máquina local. Y todo esto de manera transparente al usuario.

Para ejecutar las funciones en remoto, el plugin se conecta a una red grid que es la que lleva a cabo la ejecución de las tareas y devuelve finalmente los resultados. Como se ha explicado en secciones previas, hay diferentes tipos de redes grid, siendo Gridway la que se utilizará en este proyecto.

El acceso a una red grid necesita autenticación, y por ello el usuario del plugin debe activar la conexión desde el plugin para poder utilizar las funcionalidades que este ofrece, a través de algunos parámetros de configuración.

Cuando un usuario se conecta a la red grid tendrá las siguientes posibilidades:

- Envío de trabajos
- Monitorización de trabajos
- Cancelación de trabajos
- Obtención de resultados

Con ese objetivo conseguido tendremos un completo sistema de ejecución de funciones complejas, con alta demanda de capacidad de ejecución, a través de un acceso a red grid transparente para el usuario de la aplicación Microsoft Excel.

4.2 Casos de uso del sistema

Se van a detallar los casos de uso de las funcionalidades del Plugin Excel Grid.

Casos de uso:

1. Conexión a red Grid
2. Lanzar un trabajo a ejecución
3. Monitorizar un trabajo
4. Cancelar un trabajo
5. Ver resultados de un trabajo

4.2.1 Caso de uso “Conexión a red Grid”

Objetivo	en	Permite la conexión a la red gridway
concreto		
Precondiciones		Los campos de configuración deben estar correctamente rellenos por el usuario en la pestaña Settings del plugin Excel grid
Poscondición éxito	si	Se ha abierto una conexión con la red gridway
Poscondición fallo	si	Se informa del tipo de error que ha impedido la conexión a la red gridway
Actores		Sistema, SSH y red gridway
Secuencia normal	Paso	Acción
	1	El sistema inicializa SSH a partir de los datos de configuración del host, usuario, clave y prompt
	2	El sistema inicializa el Proxy, a partir de la clave del proxy

Secuencia alternativa	Paso	Acción
	1	No se puede inicializar la sesión SSH porque algún dato sea incorrecto o no se disponga de conexión a Internet
	2	No se puede inicializar el Proxy porque la clave del Proxy sea incorrecta

4.2.2 Caso de uso “Lanzar un trabajo a ejecución”

Objetivo	en	Permite lanzar un trabajo o array de tareas a ejecución a la red grid
Precondiciones		Ha de existir una conexión a la red gridway activa
Poscondición éxito	si	Se ha lanzado un trabajo a la red gridway para su ejecución
Poscondición fallo	si	Se informa del tipo de error que ha impedido el lanzamiento de trabajo
Actores		Usuario, sistema, SSH y red gridway
Secuencia normal	Paso	Acción
	1	El usuario se sitúa en la pestaña Job Submit

2	El usuario escribe el nombre del trabajo
3	El usuario elige el tipo de ejecutable (local o remoto)
4	El usuario especifica el ejecutable (la ruta en local o el nombre en remoto)
5	El usuario selecciona las celdas de entrada sobre las que se aplicará el ejecutable
6	El usuario pulsa el botón Submit para enviar el trabajo
7	El usuario recibe un mensaje de confirmación de envío con el ID del trabajo en la red gridway

4.2.3 Caso de uso “Monitorizar un trabajo”

Objetivo	en	Permite monitorizar el estado de un trabajo y seguir su evolución
Precondiciones		Ha de existir una conexión a la red gridway activa y se debe de haber enviado el trabajo
Poscondición	si éxito	Se ha actualizado la tabla de trabajos y se ha visualizado el trabajo y su estado actual
Poscondición	si	Se informa del tipo de error que ha impedido la

fallo	monitorización del trabajo	
Actores	Usuario, sistema, SSH y red gridway	
Secuencia normal	Paso	Acción
	1	El usuario se sitúa en la pestaña Monitoring
	2	El usuario pulsa el botón Update para actualizar la tabla de trabajos
	3	El usuario selecciona en la tabla el trabajo del que desea obtener más información
	4	El usuario pulsa el botón View Details para visualizar los detalles de dicho trabajo
Secuencia alternativa	Paso	Acción
	2	No es necesario actualizar la tabla si ya está visible
	3	No es necesario seleccionar ningún trabajo en la tabla si no se desea más información adicional de ningún trabajo (Se omitiría también el paso 4)

4.2.4 Caso de uso “Cancelar un trabajo”

Objetivo concreto	en	Permite cancelar un trabajo
Precondiciones		Ha de existir una conexión a la red gridway activa y se debe de haber enviado el trabajo
Poscondición éxito	si	Se ha cancelado el trabajo
Poscondición fallo	si	Se informa del tipo de error que ha impedido la cancelación del trabajo
Actores		Usuario, sistema, SSH y red gridway
Secuencia normal	Paso	Acción
	1	El usuario se sitúa en la pestaña Monitoring
	2	El usuario pulsa el botón Update para actualizar la tabla de trabajos
	3	El usuario selecciona en la tabla el trabajo del que desea cancelar
	4	El usuario pulsa el botón Cancel Job para cancelar dicho trabajo
	5	El trabajo se elimina de la tabla de trabajos

Secuencia alternativa	Paso	Acción
	2	No es necesario actualizar la tabla si ya está visible

4.2.5 Caso de uso “Ver resultados de un trabajo”

Objetivo en concreto Permite actualizar la hoja de cálculo con los resultados de un trabajo

Precondiciones Ha de existir una conexión a la red gridway activa y se debe de haber completado el trabajo

Poscondición si éxito Se han visualizado los resultados del trabajo

Poscondición si fallo Se informa del tipo de error que ha impedido la actualización de los resultados del trabajo

Actores Usuario, sistema, SSH y red gridway

Secuencia normal	Paso	Acción
	1	El usuario se sitúa en la pestaña Monitoring
	2	El usuario pulsa el botón Update para actualizar la tabla de trabajos
	3	El usuario selecciona en la tabla el trabajo del que desea

		visualizar los resultados (que ha de estar en estado zombie)
	4	El usuario pulsa el botón View Details para visualizar los detalles de dicho trabajo
	5	El usuario selecciona las celdas de salida sobre las que se visualizarán los resultados del trabajo
	6	El usuario pulsa el botón Copy Results para visualizar los resultados
Secuencia alternativa	Paso	Acción
	2	No es necesario actualizar la tabla si ya está visible
	3	No es necesario seleccionar el trabajo si ya están visualizados sus detalles (Se omitiría también el paso 4)

4.3 Diseño del plugin Excel Grid

Se van a detallar los requisitos y restricciones del proyecto, así como detalles de implementación del plugin.

4.3.1 Restricciones y requisitos del sistema

4.3.1.1 Descripción general

Como ya se ha comentado previamente el objetivo de este proyecto era ampliar las posibilidades de la aplicación Microsoft Excel para poder realizar cálculos complejos que requieran una gran potencia de cálculo, en otras máquinas o servidores que estarán distribuidos a lo largo de una red Grid, de manera totalmente transparente al usuario, y así liberar a la máquina local de esa carga excesiva de trabajo.

Así el usuario podrá elegir el rango de celdas de entrada sobre las que realizará el cálculo, y también posteriormente elegirá el rango de celdas de salida donde una vez completado el trabajo en la máquina remota, se visualizarán los resultados obtenidos.

Se permitirá por lo tanto elegir los datos de configuración de la red, el nombre del trabajo y la función a ejecutar.

Así mismo, además del envío de trabajos, se permitirá la cancelación de trabajos y su monitorización para conocer su estado en cualquier momento.

4.3.1.2 Contexto de uso

Perfiles de usuario o participantes

El usuario: Serán básicamente usuario de Microsoft Excel que necesiten realizar cálculos muy complejos o con una gran necesidad de potencia ya sea por el cálculo en sí o por el gran número de datos a los que se debe aplicar dicha función, de manera que no pueden ejecutarlos en su máquina local y necesitan recurrir al cómputo externo para completar sus tareas.

Otros participantes: El acceso al cómputo externo en una red grid, requiere el pacto entre el usuario y la red grid que prestará sus servicios. Es por eso que la entidad que proporciona acceso a la red grid juega su papel en este proyecto. El usuario necesita tener un certificado digital válido, y una cuenta de usuario que le darán acceso a los recursos de la red.

Usos del software

Principalmente para dar soporte a sectores de finanzas o economía, que necesitan gran potencia de cálculo para la aplicación Microsoft Excel.

Actores

Los actores que tendrán una parte representativa en la utilización del plugin son:

- El usuario del plugin
- SSH
- Globus
- Gridway

4.3.1.3 Requisitos funcionales

Conexión a una red Grid

Para todas las operaciones del plugin Excel Grid, será necesario establecer previamente una conexión a una red Grid.

Para ello, el usuario rellenará los campos del formulario correspondientes a la configuración de la red, concretamente en la pestaña Settings del plugin.

Estos datos son el nombre del host remoto, el nombre de usuario, la contraseña de usuario, la contraseña del Proxy, y el Prompt de la máquina remota.

De esta manera, cada vez que se realice una tarea, como el envío de un trabajo, la cancelación de un trabajo, la monitorización de un trabajo o su visualización de resultados, el sistema establecerá previamente dicha conexión a partir de estos datos de configuración.

La forma de realizar la conexión será por una parte la inicialización de la sesión SSH, que será la técnica utilizada para la comunicación con la máquina remota desde Windows. Para ello se utilizarán los siguientes datos de configuración: el nombre del host remoto, el nombre de usuario, la contraseña de usuario, y el Prompt de la máquina remota.

Posteriormente se inicializará el Proxy, para lo cual se utilizará la clave del Proxy introducida por el usuario.

Una vez realizadas estas acciones, el sistema estará preparado para cualquier tipo de comunicación con la red grid.

Lanzamiento de un trabajo para su ejecución

La principal finalidad del proyecto, es la posibilidad de lanzar trabajos desde la aplicación Microsoft Excel a una red Grid, para su ejecución alguna máquina remota de esa red.

Para ello, el usuario rellenará los campos del formulario correspondientes a la configuración del trabajo, concretamente en la pestaña Submit Job del plugin.

Estos datos son:

- El nombre para el trabajo
- El ejecutable: Se permitirá al usuario elegir entre un ejecutable remoto (que se encontrará en la máquina destino), del que se especificará el nombre, o bien, un ejecutable local, del que se especificará la ruta completa para su envío a la red grid.
- El rango de celdas de entrada: El usuario seleccionará el rango de celdas de entrada sobre cuyos valores se aplicará la función.

Para el lanzamiento del trabajo, primero se formarán los ficheros de entrada “input” que constituirán los parámetros de entrada de la función y el fichero plantilla o “template”.

Estos ficheros se almacenarán en un directorio que llevará el nombre del trabajo especificado por el usuario, y que se enviará a la red grid.

Se podrán tener varios parámetros de entrada para un mismo ejecutable. Así, el número de columnas especificadas en el rango de celdas de entrada, será el número de parámetros de entrada para el ejecutable, e irá en un mismo fichero “input”.

Así mismo, el número de filas existentes también jugará un papel importante, ya que el lanzamiento del trabajo será un envío de un “bulk” o array de tareas, de manera que cada fila será una de ellas, y por tanto un fichero “input” cuya extensión será el número de la tarea dentro del array.

Una vez formados estos ficheros “input”, se formará el fichero “template”, en el cual se especificará el nombre del ejecutable, el formato del nombre de los ficheros de entrada “input”, el de los ficheros de salida “output”, y el de los ficheros de error “err”.

El siguiente paso es realizar la conexión a la red grid, y llamar a la función gwsbmit especificando el número de tareas del array y el directorio donde se encuentran tanto el template, como los ficheros de entrada “input”.

Una vez realizado el lanzamiento del trabajo se le devolverá al usuario un mensaje de información con el número de ID que tiene su trabajo en la red grid, para que le sea posible la posterior monitorización y visualización de resultados.

Monitorizar un trabajo

Otra de las funcionalidades principales de la aplicación, es la posibilidad de monitorizar los trabajos enviados y poder controlar su estado en todo momento.

Para ello, existirá una tabla de trabajos, en la que se muestren todos los trabajos enviados y su estado. El usuario podrá actualizar dicha tabla en cualquier momento para seguir la evolución de sus trabajos.

Así mismo, si desea una información más precisa del estado de algún trabajo, lo podrá seleccionar en la tabla y tendrá una opción de visualizar todos los detalles de dicho trabajo.

Para llevar a cabo esta monitorización, se realiza una conexión a la red grid, y se llama a la función gwps, que devuelve el historial de trabajos lanzados, y con esos datos será con los que se rellene la tabla de trabajos.

Visualización de resultados de un trabajo

Una vez que alguno de los trabajos se encuentra en estado zombie o completado, es posible visualizar los resultados de la ejecución en la hoja de cálculo de Microsoft Excel.

Para ello, el usuario escogerá el rango de celdas de salida en la hoja de cálculo donde se colocarán los resultados.

Se comprobará que el espacio de celdas elegido corresponde con el número de tareas del array del trabajo, y una vez que todo es correcto se actualizarán dichos resultados.

Para esta actualización, se hará previamente una conexión a la red Grid, y se traerán los ficheros “output” o de resultado, así como los “err” o ficheros de error a la máquina local. Posteriormente se abrirán dichos ficheros y se sacarán los resultados a las celdas.

4.3.1.4 Requisitos estéticos

La interfaz

El diseño de la interfaz gráfica juega un papel importante en el desarrollo del proyecto, ya que será el medio que tenga el usuario para interactuar con la red Grid, y ha de ser lo más vistosa y amigable posible, de manera que el usuario se sienta cómodo y seguro al lanzar sus trabajos a través de una red.

Se intenta realizar una interfaz sencilla pero vistosa, que llame la atención al usuario para su utilización. Para ello se han utilizado iconos e imágenes que le den un aspecto luminoso.

4.3.1.5 Requisitos de utilidad

Facilidad de uso y aprendizaje

Como se comentaba anteriormente, se ha diseñado una interfaz sencilla, tanto a nivel estético como funcional. Esto es muy importante para que el usuario no tenga problemas al utilizar el plugin y pueda aprender cómo sacarle partido a todas sus funcionalidades en poco tiempo.

De esta manera, estará disponible para todo tipo de usuarios de Microsoft Excel y se le podrá sacar el máximo partido para el beneficio de los usuarios.

Para facilitar aún más el aprendizaje de la aplicación, se ha incorporado un botón de ayuda que permitirá obtener información de la utilidad de cada uno de los controles del plugin.

Personalización e internacionalización de los requisitos

Para que el plugin sea abierto a cualquier tipo de usuario, sin importar su nacionalidad o idioma, se ha realizado en inglés, lo que permite que sea estándar para cualquier persona y fácil de entender a nivel mundial.

Requisitos de accesibilidad

Para acceder al plugin será necesario contar con la aplicación Microsoft Excel, así como con un acceso a la red Grid, que le sea proporcionado por algún tipo de proveedor.

Así se deberá contar con el nombre del host remoto que de acceso a la red Grid, y una cuenta de usuario y contraseña en dicha máquina, así como de la clave del Proxy, para poder establecer las conexiones necesarias.

4.3.1.6 Requisitos operacionales

Para que el plugin pueda ser instalado y funcione correctamente es completamente necesario que el usuario cumpla con los requisitos hardware y software descritos en la siguiente sección.

Además, como se comentaba en el apartado anterior, el usuario necesita tener un certificado digital válido, y una cuenta de usuario que le darán acceso a los recursos de la red grid.

La red a la que se conecte debe proporcionar acceso ssh, y el tipo de red debe tener instalado el metaplanificador GridWay que funciona sobre Globus Toolkit.

Si no se cumplen las condiciones citadas anteriormente, el producto no funcionará o lo hará de manera incorrecta.

4.3.1.7 Requisitos de seguridad

Ya que la seguridad es un aspecto importante en cuanto entran en juego las redes y las comunicaciones, y el envío de ficheros, en nuestra aplicación se contará con la seguridad que aporta las conexiones de tipo SSH.

Todas las conexiones a la red Grid y los envíos de ficheros en ambos sentidos irán bajo la seguridad que garantiza este tipo de conexiones.

4.3.1.8 Persistencia de datos

La aplicación tiene la capacidad de almacenar los datos de configuración y la información de los trabajos enviados a la red Grid en la máquina local para poder disponer de dichos datos entre distintas sesiones si el usuario lo desea.

De esta manera aunque se cierre el plugin y la hoja de cálculo de Microsoft Excel, quedarán almacenados los datos para poder seguir trabajando con ellos cuando el usuario vuelva a arrancar la aplicación.

Por un lado, se almacenarán en un archivo “mappings” los datos de configuración: tanto los de la configuración de la conexión con la red Grid como los de la configuración del último trabajo enviado (nombre del trabajo, ejecutable, y rango de celdas de entrada).

Al guardar los datos de configuración de la conexión con la red Grid, se ofrece la posibilidad al usuario de guardarlos en su totalidad, de guardar todo excepto las contraseñas, o de no guardarlos.

Por otro lado, se almacenarán en un archivo “trabajos” la información referente a los trabajos enviados a la red Grid. Se almacenan tanto los trabajos que se encuentran en estado pendiente como los que ya han sido completados pero aún no se ha visualizado el resultado. Una vez que se visualizan los resultados de un trabajo en la hoja de cálculo, éste se borra del fichero.

La información que se guarda de cada uno de los trabajos será su ID en la red Grid, su nombre, el número de tareas totales, el número de tareas completadas hasta el momento y la fecha de lanzamiento.

De esta manera cada vez que se arranca la aplicación, se cargarán dichos ficheros de datos rellenando los campos oportunos en el formulario. Del mismo modo, el fichero de trabajos estará disponible para cuando el usuario quiera realizar una monitorización de los trabajos.

4.3.1.9 Restricciones y limitaciones

Límites del proyecto

El plugin que implementamos en el proyecto solamente opera sobre equipos x86 con Windows como sistema operativo. Además es válido únicamente para el gestor de hojas de cálculo Microsoft Excel.

La red grid con la que conecta para el envío de tareas debe tener GridWay.

Requisitos Hardware

El usuario

- Pentium II o equivalente
- 64 MB de memoria RAM
- Conexión a Internet

La red grid

- No afecta al plugin.

Requisitos Software

El usuario

- Microsoft Excel
- Sistema operativo Microsoft Windows

La red grid

- ssh
- Globus
- GridWay

4.4 Entorno de implementación

Se van a detallar las tecnologías y herramientas utilizadas para el desarrollo del plugin, así como el medio físico en el que se instala nuestra aplicación.

4.4.1 Tecnologías y herramientas de desarrollo

4.4.1.1 Tecnología SSH

SSH (Secure SHell) es el nombre de un [protocolo](#) y del [programa](#) que lo implementa, y sirve para acceder a máquinas remotas a través de una red. Permite manejar por completo el [ordenador](#) mediante un [intérprete de comandos](#), y también puede redirigir el tráfico de [X](#) para poder ejecutar programas gráficos si tenemos un [Servidor X](#) arrancado.

Además de la conexión a otras máquinas, SSH nos permite copiar datos de forma segura (tanto ficheros sueltos como simular sesiones [FTP](#) cifradas), gestionar [claves RSA](#) para no escribir claves al conectar a las máquinas y pasar los datos de cualquier otra aplicación por un canal seguro [tunelizado](#) mediante SSH.

SSH trabaja de forma similar a como se hace con [telnet](#). La diferencia principal es que SSH usa técnicas de cifrado que hacen que la información que viaja por el medio de comunicación vaya de manera no legible y ninguna tercera persona pueda descubrir el usuario y contraseña de la conexión ni lo que se escribe durante toda la sesión.

4.4.1.2 Tecnología Grid

La tecnología Grid surge como un nuevo paradigma de computación distribuida a mediados de los 90. Se basa fundamentalmente en el acceso remoto a recursos computacionales, y su tecnología estándar es el Globus Toolkit. Su objetivo es permitir gestionar y distribuir la potencia de cálculo disponible, de tal forma que los usuarios se beneficien de la potencia de ordenadores infrautilizados que se encuentran dispersos geográficamente. Éstos ceden parte de sus recursos para procesar aplicaciones con funciones de cálculo avanzadas.

Gracias a la tecnología Grid, el usuario puede utilizar toda la red o redes de ordenadores para solicitar la potencia que necesita de otros equipos que no trabajan al máximo de sus posibilidades para que les preste esos recursos sobrantes. De esta manera el usuario podrá acceder a un supercomputador virtual con la potencia suficiente para realizar de forma sencilla los cálculos avanzados. Es decir, si el usuario necesita más memoria RAM y más disco

duro para almacenar un trabajo no hace falta que lo haga en el de su propio equipo, sino que puede utilizar el de los otros ordenadores de la red Grid.

Grid se diferencia de los sistemas cliente-servidor y otras tecnologías actuales (CORBA, EJB o .NET), en que está orientada a los recursos computacionales y no a la información, la seguridad no está en un segundo plano y la comunicación es asíncrona.

Grid es un paso más allá de Internet, pues incorpora gran ancho de banda, alta velocidad de procesado, y bases de datos de gran tamaño con recursos disponibles para la industria, los científicos y los ciudadanos en general. La interconexión entre ordenadores usando la tecnología Grid permite a un número elevado de usuarios obtener información sin determinar a qué ordenador quieren conectarse. Las aplicaciones que se beneficiarán con esta tecnología cubren las áreas de la Salud, la Astronomía, la Astrofísica, el Análisis Financiero, etcétera.

4.4.1.3 Microsoft Visual Studio .NET

Es el entorno de programación utilizado para el desarrollo del proyecto. Soporta los nuevos lenguajes [.NET](#): [C#](#), [Visual Basic .NET](#) y [Managed C++](#), además de [C++](#). Visual Studio .NET puede utilizarse para construir aplicaciones dirigidas a Windows (utilizando Windows Forms), Web (usando [ASP.NET](#) y [Servicios Web](#)) y dispositivos portátiles (utilizando [.NET Compact Framework](#)).

En la implementación de este proyecto nos hemos beneficiado de algunas de las ventajas de esta herramienta, como la posibilidad de:

- Desarrollar rápidamente poderosas aplicaciones. Con un ambiente de desarrollo integrado (IDE) para todos los lenguajes, los desarrolladores pueden tomar ventaja de una caja de herramientas común (toolbox), depurador (debugger) y lista de tareas, reduciendo en gran manera la curva de aprendizaje del desarrollador.
- Crear aplicaciones que funcionan en cualquier dispositivo y se integran en cualquier plataforma. Visual Studio .NET Professional proporciona a los desarrolladores las herramientas para integrar soluciones entre sistemas operativos y lenguajes.

En el desarrollo del plugin se utilizan:

- Visual Basic: Para los módulos que implementan el formulario del plugin y la conexión con Microsoft Excel.
- Visual C#: Para el módulo de conexión con la red Grid.

Debido a la interoperabilidad multilenguaje que proporciona la plataforma .NET, ha sido muy fácil integrar dichos módulos. Así mismo, ha sido muy sencillo e intuitivo el diseñar la interfaz gráfica a través de las posibilidades que ofrece Visual Basic.

4.4.2 Medio físico

Para el desarrollo del proyecto utilizamos equipos propios con el entorno de programación Microsoft Visual Studio .NET.

Además es necesaria la conexión con un servidor de red grid contra la que probar el producto. El departamento de Arquitectura y Tecnología de Computadores de la UCM nos proporciona una cuenta de usuario con certificado digital válido para operar en su red grid.

4.4.2 Software implicado en el desarrollo

En el desarrollo de nuestro software se utiliza Microsoft Excel, tanto en su testeo como en su implementación.

Lo usaremos para comprobar el funcionamiento de la aplicación. Es la aplicación de hojas de cálculo para la que desarrollamos el plugin.

5 CONCLUSIONES Y TRABAJO FUTURO

5.1 Conclusiones

5.1.1 Situación final del proyecto

Toda la funcionalidad prevista en la especificación de requisitos del Plugin Excel Grid ha podido ser desarrollada. Esta funcionalidad, adecuadamente probada, es la siguiente:

- Conexión a una red Grid a través de la tecnología SSH.
- Lanzamiento de trabajos a la red Grid.
- Monitorización de trabajos.
- Cancelación de trabajos.
- Visualización de los resultados de un trabajo.

También destacar que se permite almacenar los datos de configuración entre versiones.

5.1.2 Trabajo realizado y conocimientos adquiridos

Para llevar a cabo el desarrollo de esta herramienta, ha sido necesario realizar un trabajo coordinado y organizado, para el cual han sido muy útiles los conceptos adquiridos en la asignatura de Ingeniería del Software.

Se realizó en un principio la especificación de requisitos, así como unas metas de alcance del proyecto. Posteriormente, se decidió el tipo de arquitectura que modularía mejor la aplicación, y su división en distintos paquetes separados según la funcionalidad, para facilitar sus posteriores ampliaciones si existiesen.

En la gestión de configuración por parte de los miembros del grupo ha sido determinante el uso de internet, pues nos hemos podido comunicar día a día mediante e-mails y mediante programas de chat, para trabajar on-line.

Finalmente cada integrante del equipo de desarrollo ha cumplido satisfactoriamente con el trabajo que tenía asignado.

Los conocimientos adquiridos se centran entorno a la tecnología Grid, un nuevo paradigma de computación distribuida, que poco a poco está suponiendo y supondrá una revolución en el tipo de servicios prestados a través de Internet, gracias al enorme potencial de intercambio y de gestión de recursos que ofrece.

También hemos profundizado en el entorno de programación Microsoft Visual Studio .NET, y en los lenguajes Visual Basic y Visual C#.

Así mismo, aprendimos un nuevo tipo de arquitectura, completamente innovador y muy interesante por su actualidad: los plugins. Y como realizar uno de ellos para ser utilizado como una extensión de una aplicación de Microsoft Office, en concreto, Microsoft Excel.

5.2 Trabajo futuro

5.2.1 Ampliar la aplicación para que sirva para otro tipo metaplanificadores

Sería interesante, como futura mejora, que el plugin sirviera para otro tipo de metaplanificadores, sin estar limitado a gridway.

5.2.2 Crear un modo interactivo para la aplicación

Como ampliación de la aplicación, sería interesante que además del modo de ejecución batch, se realizara un modo interactivo, en el cual se visualizaran los resultados según se fuera completando subtarefas de cada trabajo, sin necesidad de que haya terminado en su totalidad. De este modo se podrían ir realizando también gráficas de la evolución de los resultados de forma dinámica.

Apéndice A - Manual de usuario de Plugin Excel Grid

A.1 Introducción

Plugin Excel Grid es un complemento para Microsoft Excel que permite enviar trabajos complejos o que requieran gran capacidad de cómputo a una red Grid para su ejecución, liberando a la máquina local de dicho trabajo, y pudiendo monitorizar los trabajos y visualizar sus resultados sobre la misma hoja de cálculo, de forma totalmente transparente al usuario.

Plugin Excel Grid es una herramienta de gran valor para distintos tipos de usuarios, incluyendo importantes comunidades financieras que necesiten realizar cálculos muy complejos sobre Microsoft Excel.

Plugin Excel Grid proporciona las siguientes funcionalidades:

- Conexión a una red Grid
- Envío de trabajos a la red Grid
- Monitorización de trabajos
- Cancelación de trabajos
- Visualización de resultados de los trabajos

A.2 Plugin Excel Grid

Plugin Excel Grid es una aplicación realizada en Visual Studio .Net, que constituye una extensión para Microsoft Office, por lo tanto, se puede ejecutar sobre la plataforma Microsoft Windows, que cuente con Microsoft Excel y con conexión a Internet.

Se proporciona un instalador para Microsoft Windows, que generará el directorio de instalación, incluirá todas las librerías necesarias para la instalación y ejecución de la aplicación, y lo instalará de forma que ya se pueda ejecutar de forma simple mediante un acceso directo.

A.2.1 Instalación y ejecución de Plugin Excel Grid

En el CD que se proporciona con esta memoria, hay una copia del instalador de la aplicación Plugin Excel Grid.

Para instalar la aplicación

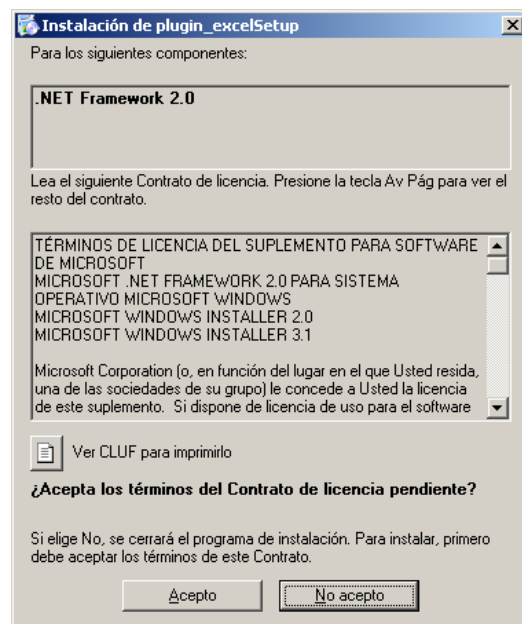
La instalación de la aplicación se realiza mediante un fichero de instalación que viene acompañado de un paquete de instalación MSI.

1º Haciendo doble clic sobre el fichero setup.exe se inicia la instalación.

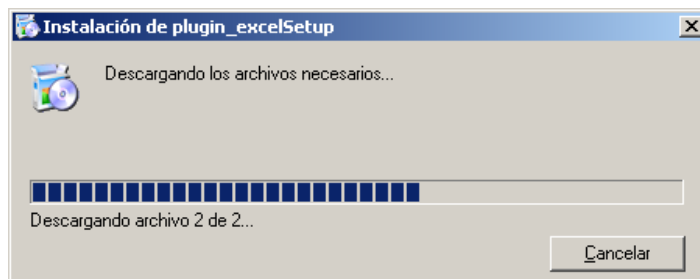
Si el equipo no tiene instalada la versión 2.0 de la plataforma .NET Framework, el asistente solicitará confirmación de instalación y aceptación de la licencia de este software para comenzar con la instalación propiamente dicha de Excel GridWay. Si el equipo ya cuenta con .NET Framework 2.0, el asistente pasará directamente al paso número 3.

2º Confirmación de instalación y contrato de licencia de la plataforma .NET Framework 2.0

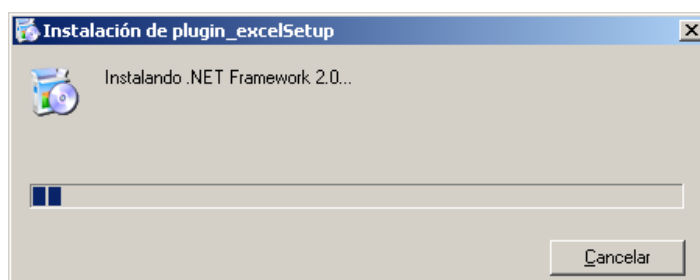
Para continuar con el proceso de instalación pulsamos el botón Aceptar.



Nos aparecerá entonces una ventana mostrando el proceso de descarga de .NET 2.0

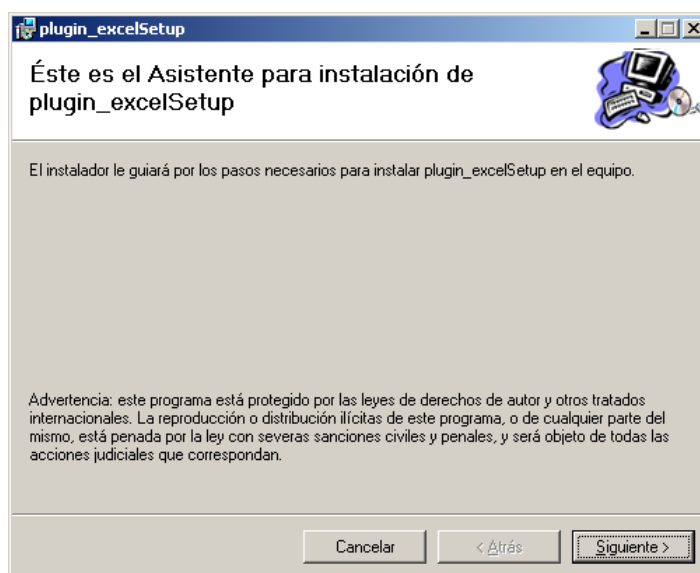


Seguidamente a la descarga veremos cómo se instala .NET 2.0 en nuestro equipo.



3º Inicio del proceso de instalación de Excel GridWay

Pulsamos Siguiente para continuar con la instalación:



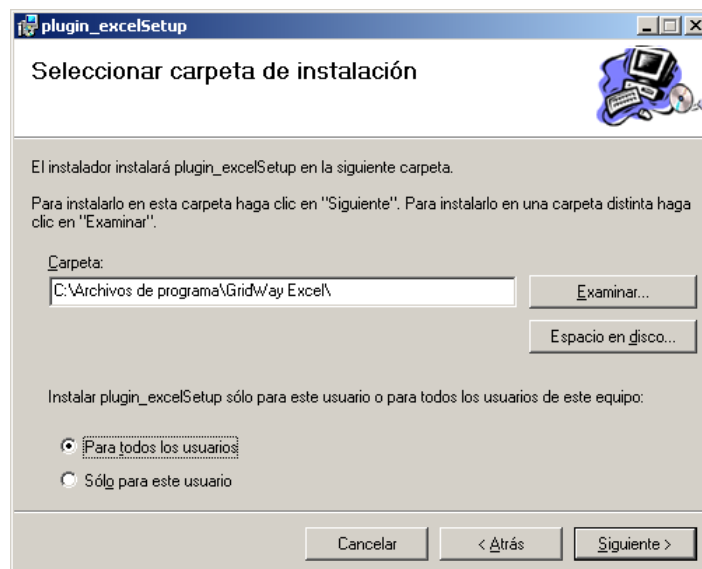
4º Selección de la carpeta de destino de los ficheros de instalación.

De forma predeterminada viene escrita una ruta que indica dónde se instalará el programa. Podemos cambiarla si preferimos otro lugar.

En esta pantalla se puede elegir qué usuarios del sistema tendrán disponible la aplicación una vez esté instalada (opción no disponible en versiones Windows Me, 98). Por defecto aparecerá

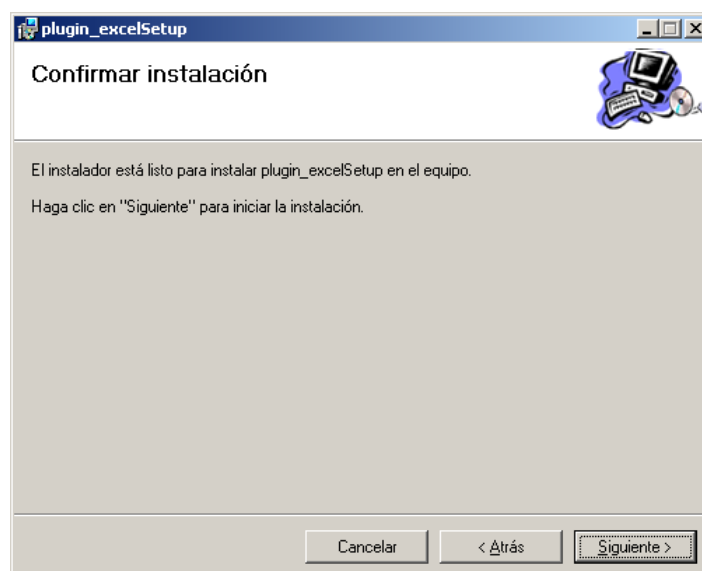
marcada la opción “Para todos los usuarios”, pero puede elegir “Sólo para este usuario” si así lo prefiere.

Hechas las elecciones podemos continuar pulsando Siguiente.



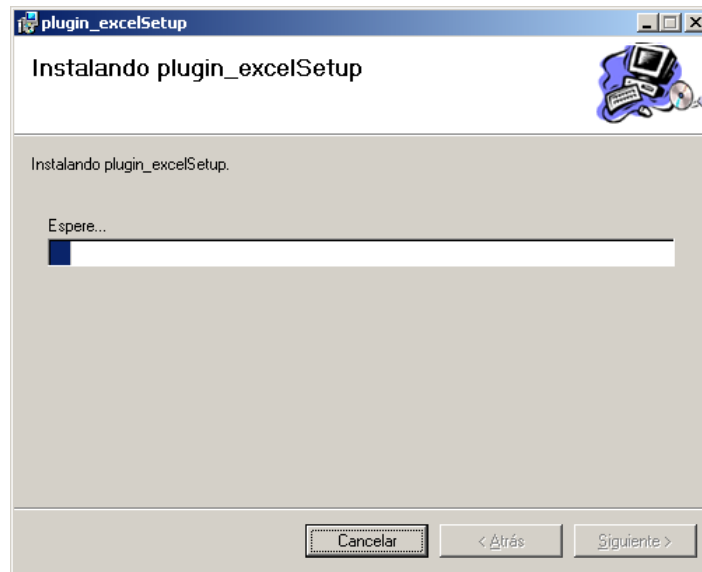
5º Confirmación de la instalación.

Confirmamos finalmente que efectivamente deseamos realizar la instalación pulsando Siguiente.



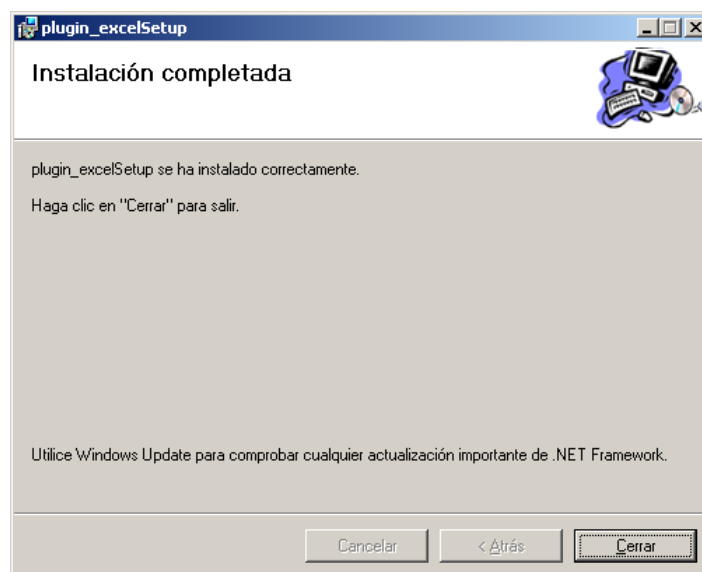
6º Copia de ficheros de instalación.

El asistente de la instalación comienza la copia de ficheros de la instalación. Este proceso se muestra con una barra de progreso que representa el porcentaje de ficheros copiados.



7º Fin de la instalación.

El asistente nos indica que la aplicación se ha instalado correctamente y listo para utilizarse. Cerramos la ventana pulsando Cerrar.



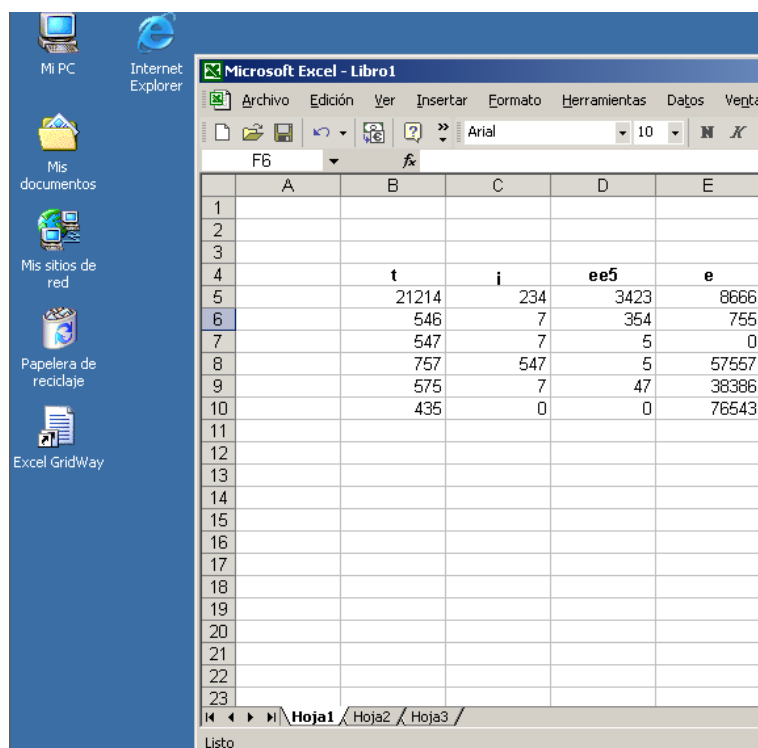
Para ejecutar la aplicación

Para utilizar la aplicación e interactuar con Excel debemos tener éste último abierto.

Para abrir la aplicación Excel GridWay tenemos dos opciones:

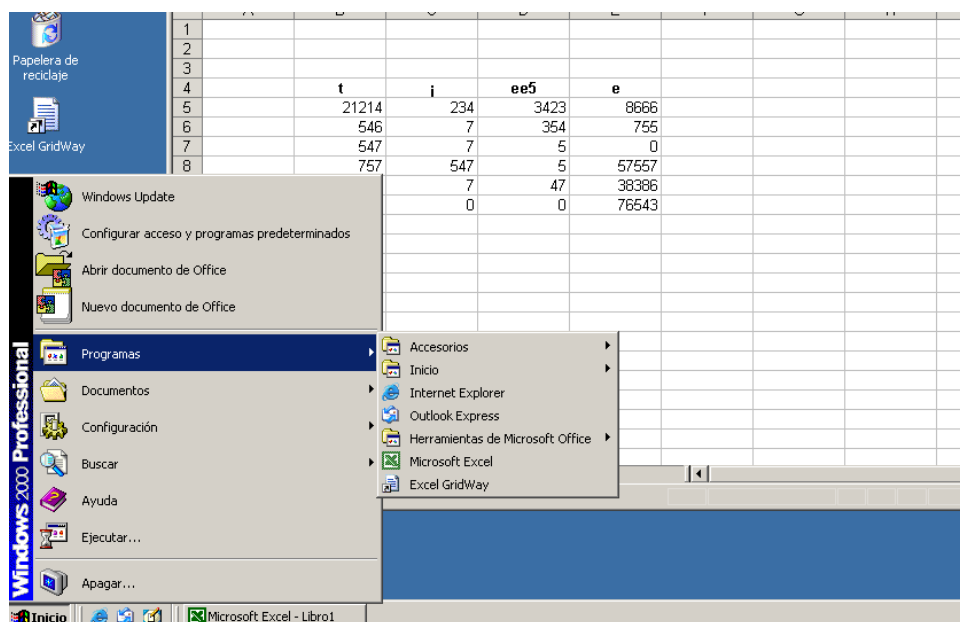
Desde el escritorio

El programa de instalación deja un acceso directo en el escritorio. Haciendo doble clic sobre él abriremos la aplicación.



Desde el menú inicio

Accediendo al menú Inicio → Programas → Excel GridWay abriremos la aplicación.



A.2.2 Requerimientos del sistema

En cuanto a hardware, los requerimientos mínimos del sistema son:

- Pentium II o equivalente
- 64 MB de memoria RAM
- Conexión a Internet

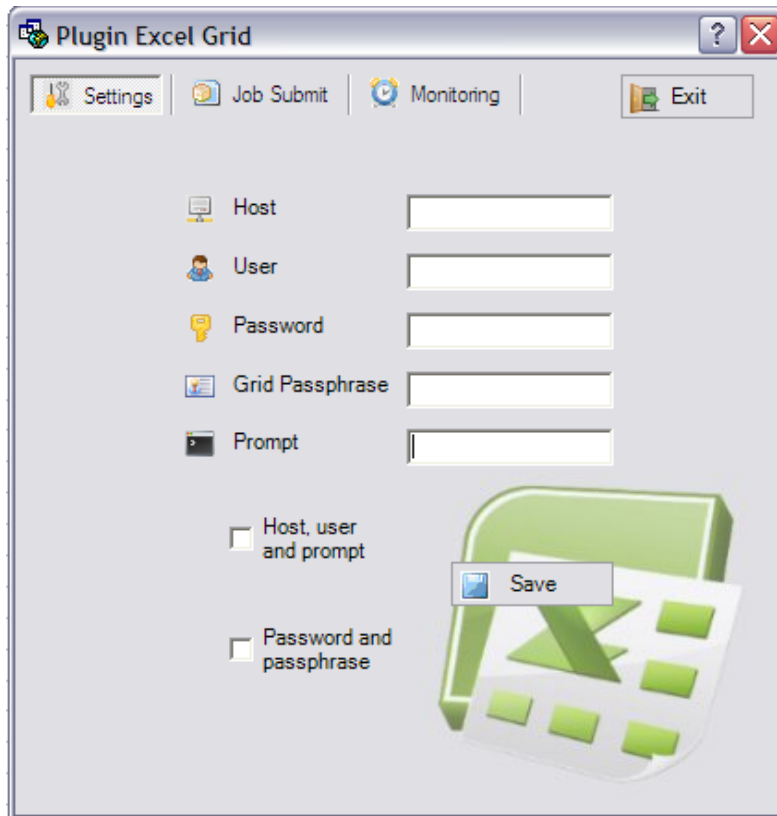
En cuanto a software:

- Microsoft Windows 98, ME, NT, 2000, o XP
- Microsoft Excel XP o superior

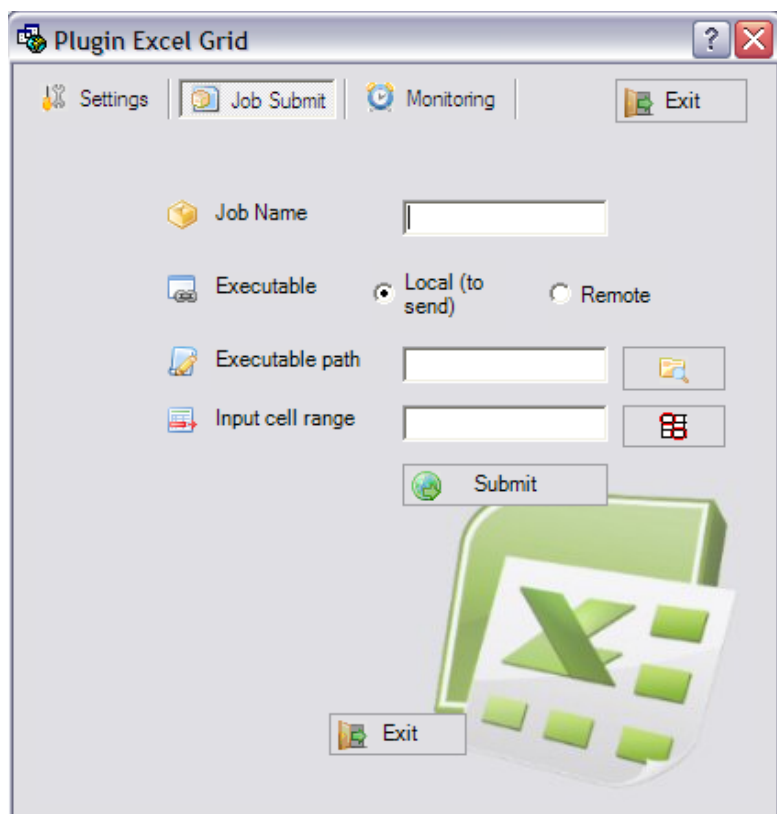
A.2.3 Elementos del Plugin Excel Grid

La aplicación se compone básicamente de tres pestañas:

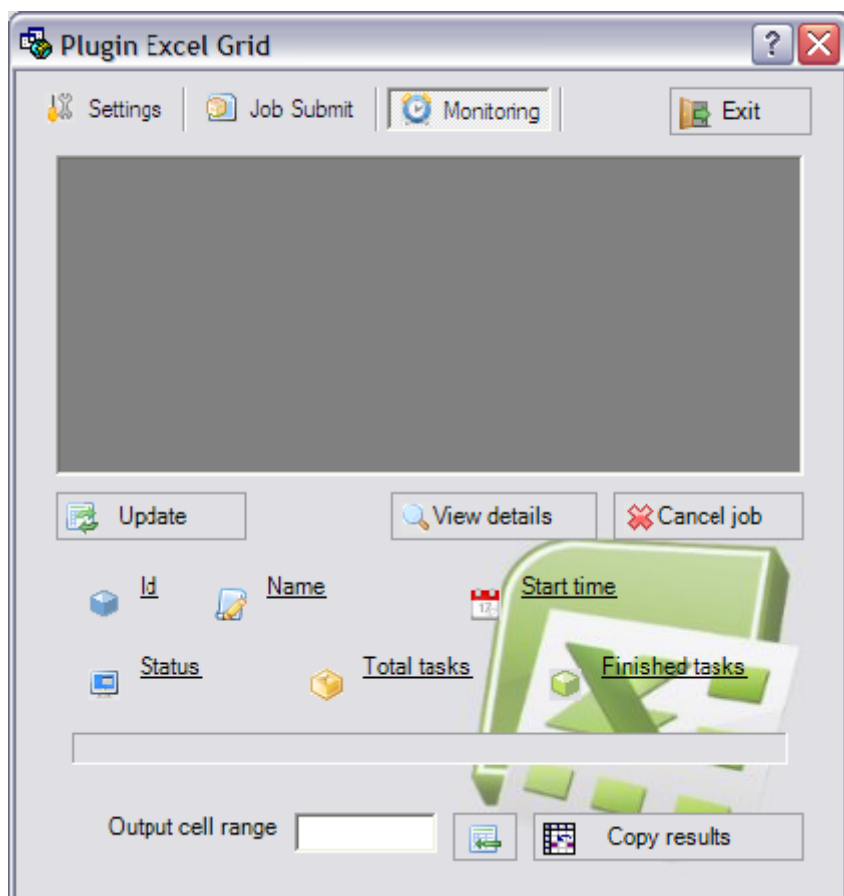
- Settings, que permitirá configurar la conexión a la red Grid.
- Job Submit, que permitirá configurar un trabajo y enviarlo a la red Grid.
- Monitoring, que permitirá monitorizar los trabajos previamente enviados a la red Grid, cancelar algún trabajo si se desea, y visualizar los resultados de los trabajos completados sobre la hoja de cálculo.



Pestaña Settings



Pestaña Job Submit



Pestaña Monitoring

A.3.Tutorial

Esta sección explica cómo configurar la conexión a la red Grid, cómo lanzar nuevos trabajos a la red, como monitorizarlos, cancelarlos y visualizar sus resultados en la hoja de cálculo una vez completados.

A.3.1 Configuración de la conexión a la red Grid:

Para realizar el envío de un nuevo trabajo a la red Grid, o bien la monitorización, cancelación o visualización de resultados de los trabajos ya enviados previamente, es necesario haber configurado previamente la conexión a la red Grid, ya que se tendrá que conectar a dicha red para ejecutar el resto de las funcionalidades.

De este modo, para configurar la conexión, el usuario deberá lanzar la aplicación y situarse en la pestaña Settings.

Una vez en esta pestaña, se deberán completar todos los campos del formulario, con los datos de su cuenta de usuario que le posibilita su conexión a la red Grid.

Así, deberá completar los siguientes campos:

- **Host:** Nombre de la máquina a la cual se conectará de forma remota y en la que se tiene la cuenta de usuario que posibilitará el acceso a la red Grid.
- **User:** Nombre del usuario referente a la cuenta de la que se cuenta en el host.
- **Password:** Contraseña de la cuenta de usuario.
- **Grid Passphrase:** Contraseña del Proxy.
- **Prompt:** Prompt del host.

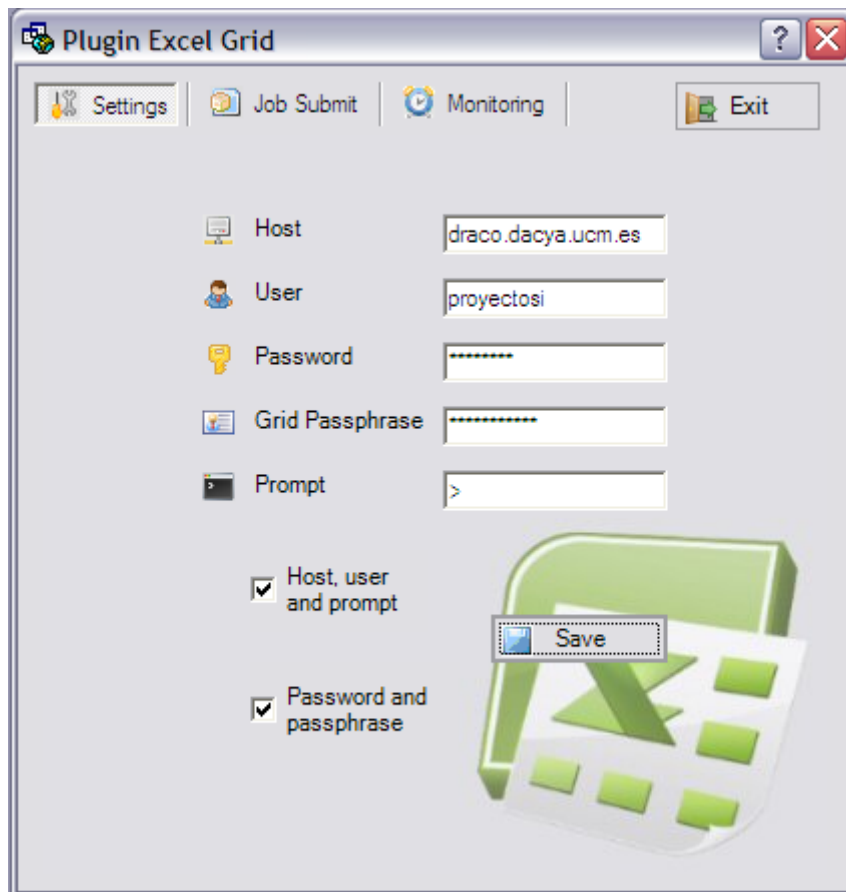
Por otro lado, se tienen dos casillas que sirven para especificar si se desea o no guardar los datos de configuración que se acaban de introducir para que estén disponibles en próximas sesiones.

Así, si se señala la primera casilla “Host, user and prompt”, estamos especificando que deseamos que los valores de estos tres campos se guarden para la próxima sesión.

Del mismo modo, si también señalamos la casilla “Password and passphrase”, se guardarán también las dos contraseñas introducidas para la siguiente sesión.

Finalmente, una vez rellenados los campos, si hay marcada alguna de las dos casillas, el usuario ha de pulsar el botón Guardar, para que realmente se efectúe el almacenamiento en el correspondiente fichero de datos.

A continuación se muestra un ejemplo de cómo se podría rellenar dicha pestaña:



Ejemplo Pestaña Settings

A.3.2 Crear y lanzar un nuevo trabajo a la red Grid:


Para realizar el envío de un nuevo trabajo a la red Grid, el usuario se debe colocar en la pestaña Job Submit (previamente debe haber configurado la conexión a la red Grid, paso A.3.1).

Una vez situado en la pestaña, deberá completar los campos del formulario:

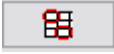
Job Name: Nombre que se le desea dar al trabajo, será el nombre del directorio que se creará para almacenar los distintos ficheros de configuración, de datos de entrada y de resultados, tanto en local, como en el host remoto.

Executable: Se deberá elegir una de entre las dos opciones de ejecutable. La primera, “Local” permitirá especificar la ruta completa del ejecutable, que ha de estar en la máquina local, de manera que este fichero se enviará a la máquina remota. La otra opción es elegir la

opción “Remote”, siendo el usuario el que especifique el nombre del ejecutable que se debe encontrar en la máquina remota de ejecución.

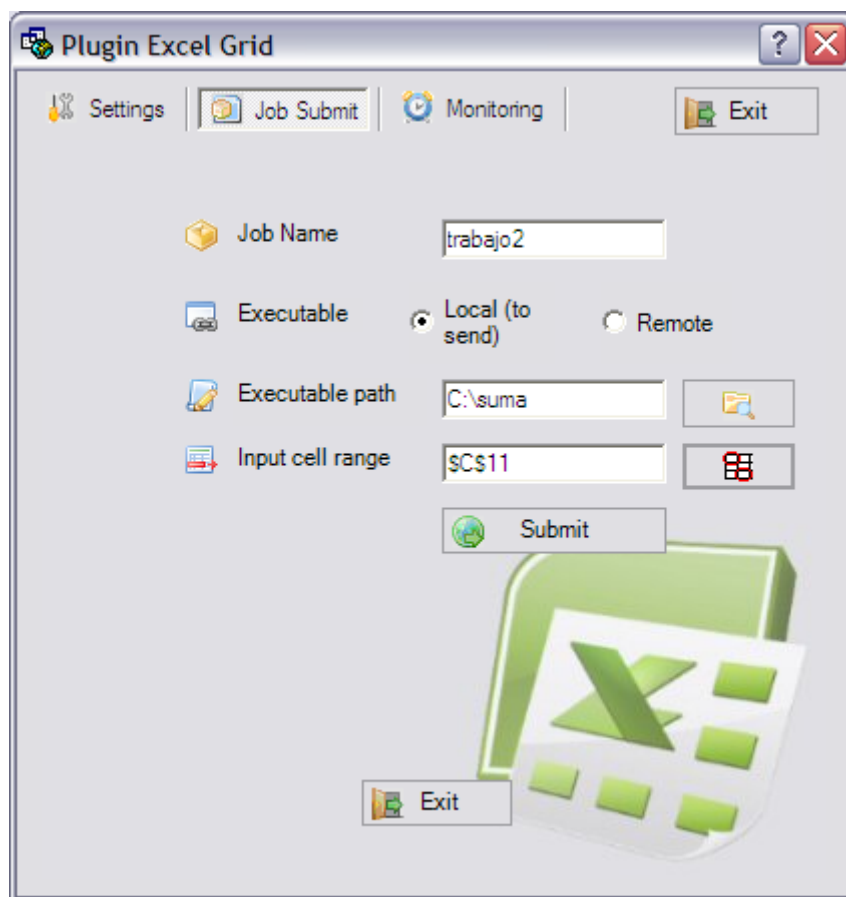
Executable Path: Este campo aparecerá cuando se elige como tipo de ejecutable “Local”. El usuario elegirá la ruta completa donde se encuentra dicho ejecutable pulsando el botón que aparece a la derecha . Se abrirá un explorador de archivos, y se recorrerá el árbol de archivos hasta encontrar el ejecutable. La ruta se rellenará de forma automática en el campo de edición.

Executable File: Este campo aparecerá cuando se elige como tipo de ejecutable “Remote”. El usuario especificará el nombre del ejecutable remoto.

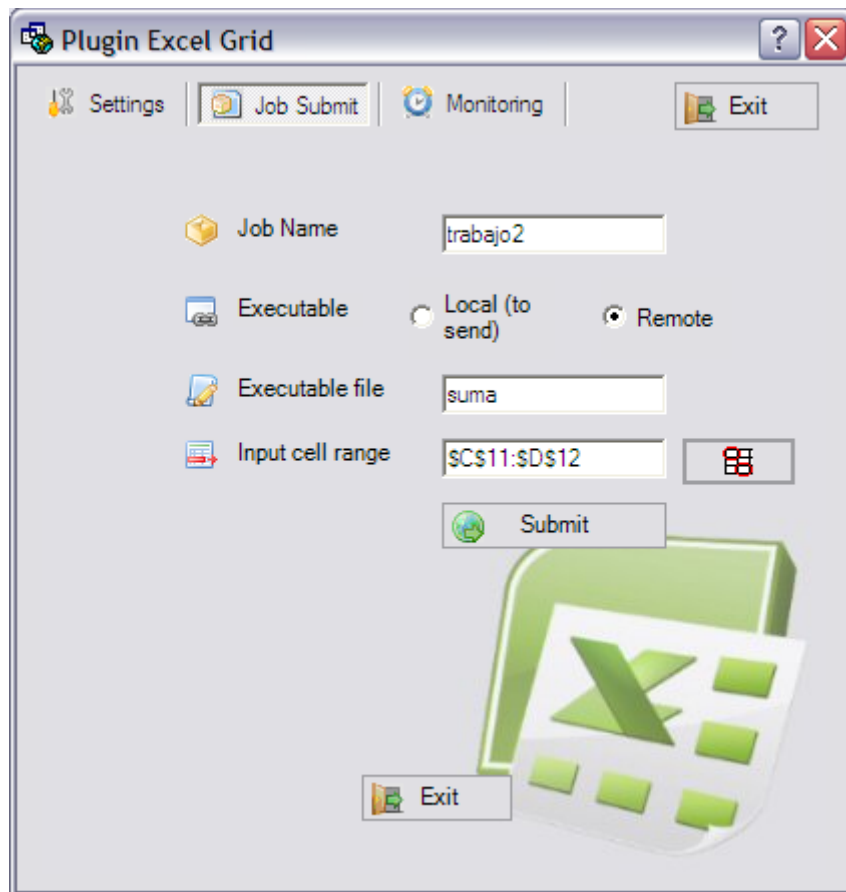
Input Cell Range: Este campo se rellenará de forma automática cuando el usuario una vez pulsado el botón que aparece a la derecha , seleccione sobre la hoja de cálculo el rango de celdas que desea que sean los datos de entrada para el cálculo de su trabajo.

Por último, una vez que están completos todos los campos, el usuario deberá pulsar el botón Submit, para que se realice el envío del trabajo a la red Grid.

Un ejemplo de cómo podría quedar completada dicha pestaña sería:



Ejemplo Pestaña Job Submit, ejecutable local



Ejemplo Pestaña Job Submit, ejecutable remoto

A.3.3 Monitorizar un trabajo:

Para realizar la monitorización de un trabajo, el usuario se debe colocar en la pestaña Monitoring (previamente debe haber configurado la conexión a la red Grid, paso A.3.1, y haber enviado el trabajo que desea monitorizar, paso A.3.2).

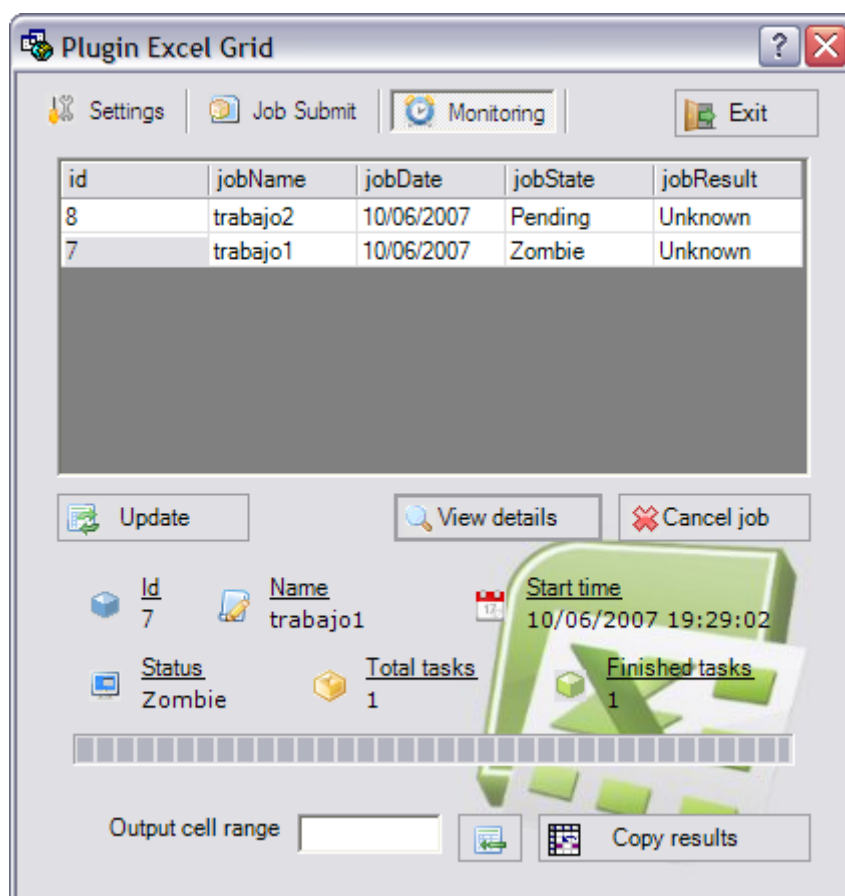
Una vez situado en la pestaña, el usuario deberá pulsar el botón Update para actualizar la tabla con los trabajos del usuario que se encuentran actualmente en la red Grid.

Sobre la tabla, el usuario deberá seleccionar con el ratón la fila del trabajo del cual desea visualizar más detalles, y pulsar el botón View Details.

Automáticamente se visualizarán en los campos de la parte de debajo de la pestaña, todos los datos referentes al estado del trabajo.

Si el trabajo se encuentra en estado “Zombie”, quiere decir que está completado, y si está “Pending”, quiere decir que aún no se han completado todas sus subtareas. En la barra de progreso se verá el porcentaje de subtareas completadas.

Un ejemplo de la monitorización puede ser el siguiente:



Ejemplo Pestaña Monitoring (Visualización de un trabajo)

A.3.4 Cancelar un trabajo:

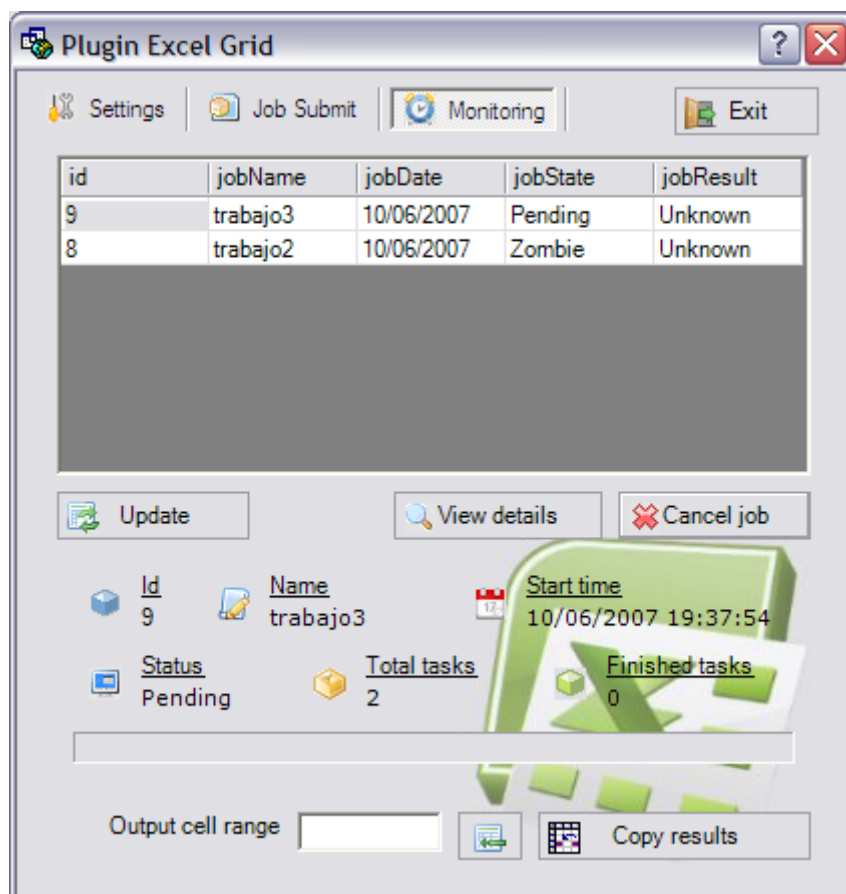
Para cancelar un trabajo, el usuario se debe colocar en la pestaña Monitoring (previamente debe haber configurado la conexión a la red Grid, paso A.3.1, y haber enviado el trabajo que desea cancelar, paso A.3.2).

Al pulsar el botón Update, se actualizará la tabla de trabajos.

El usuario deberá seleccionar con el ratón el trabajo que desea cancelar y pulsar el botón View Details.

Se mostrarán los detalles del trabajo seleccionado. Ahora es el momento de pulsar el botón Cancel Job, de manera que el trabajo se elimina de la red grid, y de la lista de trabajos.

A continuación se muestra un ejemplo de la cancelación de un trabajo:



Ejemplo Pestaña Monitoring (Cancelación de un trabajo)


A.3.5 Visualizar los resultados de un trabajo:

Para visualizar los resultados de un trabajo, el usuario se debe colocar en la pestaña Monitoring (previamente debe haber configurado la conexión a la red Grid, paso A.3.1, y haber enviado el trabajo que desea cancelar, paso A.3.2).

Se deberán pulsar los botones Update para actualizar la tabla de trabajos y View Details habiendo seleccionado el trabajo del que se desean visualizar los resultados, al igual que se hace en la monitorización (paso A.3.3).

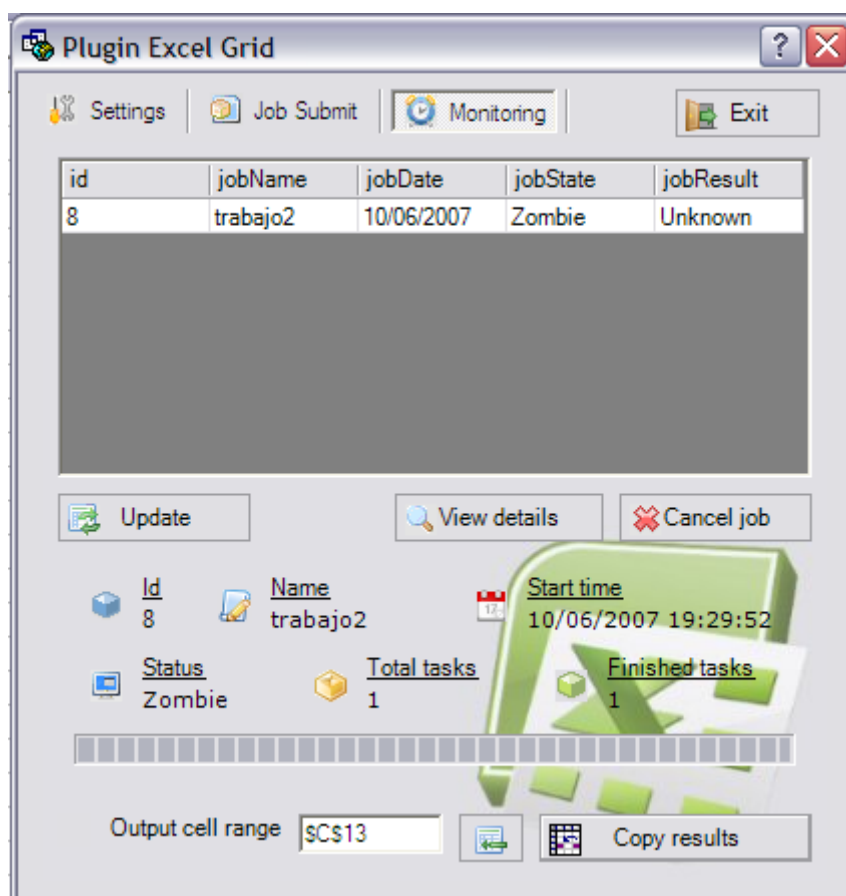
Una vez visualizado el trabajo, y estando éste en estado “Zombie”, es decir, comprobando que se ha finalizado por completo y que el “Total tasks” es igual al “Finished tasks” y que la

barra de progreso está completa, el usuario ha de seleccionar el rango de celdas de salida donde se quieren visualizar los resultados.

Para ello, el campo “Output Cell Range” se rellenará de forma automática cuando el usuario una vez pulsado el botón que aparece a la derecha , seleccione sobre la hoja de cálculo el rango de celdas donde los desea visualizar (deben ser tantas celdas como total de subtareas del trabajo) .

Finalmente, el usuario deberá pulsar el botón Copy Results, y los resultados se copiarán sobre las celdas seleccionadas automáticamente.

Un ejemplo de esta visualización de resultados sería:



Ejemplo Pestaña Monitoring (Visualización de resultados)

Apéndice B - HowTo Inglés

B.1 Plugin Excel

The application is composed for three packages o modules.

The first module is called “plugin-excel”.

This package is based on the construction of an add-in for Microsoft Excel using Microsoft Visual Basic .NET. These add-ins are called COM add-ins and Microsoft Office 2000 and later support an uniform design architecture for building these application add-ins to enhance and to control Office applications.

The way to build these type of add-ins is described step-by-step in this article, as well as the most important subroutines on this module.

A COM add-in is an in-process COM server, or ActiveX dynamic link library (DLL), that implements the IDTExtensibility2 interface as described in the Microsoft Add-in Designer type library (Msaddndr.dll). All COM add-ins inherit from this interface and must implement each of its five methods.

In this way, this package contains the class called Connect.vb, that implements these methods, that are:

OnConnection

The event fires whenever the COM add-in is connected. If returns successfully, the add-in is said to be loaded. If an error message is returned, the host application immediately releases its reference to the add-in, and the object is destroyed.

Takes the following four parameters:

- Application - A reference to the host application object.
- ConnectMode - A constant that specifies how the add-in is connected.
- AddInInst - A reference to the COMAddIn object that refers to this add-in in the COMAddIns collection for the host application.
- Custom - An array of Variant type values that can hold user-defined data.

We use the Application and AddInInst parameters for inicializing the Excel Application.

OnDisconnection

The event fires when the COM add-in is disconnected and just before it unloads from memory. The add-in performs a cleanup of resources in this event.

Takes the following two parameters:

RemoveMode - A constant that specifies how the add-in was disconnected.

Custom - An array of Variant type values that can hold user-defined data.

OnAddInsUpdate

The event fires when the set of registered COM add-ins changes. In other words, whenever a COM add-in is installed or removed from the host application, this event fires.

OnStartupComplete

The event is called if the add-in was connected during startup. Its functionality is create a menu command on the "Tools" menu with init parameters (as name and visibility), making sure that menu command doesn't already exist.

OnBeginShutdown

The event is called if the host disconnects the add-in during shutdown.

O_cmd_Click

The event is fired when the user push the button on the tool menu of Microsoft Excel, "plugin excel grid". Its functionality is connect the add-in with the form that is contained in the next package, plugin_excel_grid_runner, and call this package to load this form.

To build a plugin like this you must create a new Project in Microsoft Visual Studio .NET, and in Other Projects, select Extensibility Projects, and then select the Shared Add-in template. You must give a name for your add-in, select create the plugin using Visual Basic, and then select the application of Microsoft for what you would like to build the add-in. And following these steps you will get the Connect class detailed previously.

B.2 Plugin Excel Grid Runner

The main functionalities of this package are two.

- Showing a friendly and useful interface to the user

- Implements all the calling methods that bring the interface the possibility of working with Grid networks suitable functions with the assistance of the GridwayBridge package

This package is a mere Windows Application and all its composing classes are Visual Basic classes.

It contains two special references:

- Excel interop library (Microsoft.Office.interop.Excel), which allow the package and its methods to interact with the Microsoft Excel application.
- GridwayBridge package, in order to work with Grid functionalities

Next we'll explain each class working way

GSettings

This class defines the needed methods to handle the information we will use at the moment of sending a new job. It works both with specifically job information and with user details too.

It uses serialization, a Visual Basic process that converts an object into a linear sequence of bytes. It is very useful because permits to save space and time creating more efficient data access and storing.

The class implements the get and set methods of the data fields.

GTrabajos

Like GSettings, another class which uses serialization and provide us a better treatment of information. Now the information contained are specifics details of the job that will be stored in our computer in order to recover them when we ask for the state of the sent jobs.

Mappings

Thanks to this class it can be created a '.dat' file which store all data managed by GSettings and GTrabajos about configuration of the application and the sended jobs.

mainClass

This class manages the throwing of the application when we call it from Microsoft Excel with the specific option

frmGClient.vb

This is a common Visual Basic class that use the package 'System.Windows.Form' in order to build the graphic interface. All the methods and components which let our class to create a GUI are contained in that package. This class will use GSettings, Gtrabajos and mainClass classes to manage the information that is being handled.

The user interface contains three different sections that comprise diverse aspects:

- **Settings:** The user parameters are reflected here. These parameters are necessary to generate new jobs or consulting existing jobs details. They can be also kept in our system thanks to the Mappings class (GuardarConfig_Click method).
- **Job Submit:** The user can specify and send a new job from this section, assigning it a new name, the file that will be executed and the input parameters for that file, taken from a selection of Excel cells. We can choose to send the executable file from our computer or use another one guest in a remote machine (BrowsePathButton_Click method).
- **Monitoring:** Data about the user sent (and not gather) jobs is showed in this section. Here we can find a DataTable which contains all that information: job id, name, state, etc (UpdateButton_Click method). We also can focus into a specific job seeing more details as total and finised tasks or the sending date (BotonTemporal_Click method). In order to show these details we pick up the information from the GridwayBride module (gwps) and from one file updated each time we send a new job, generated with the help of GTrabajos and Mappings classes.

The updating of an Excel sheet with the results of a finished job is another of this section functions (ClearButton_Click method), getting them from the remote machina through GridwayBridge package. It's also allowed to cancel a sent job while we haven't recognized and brought it to our computer (CancelJobButton_Click method).

It also contains the functions that allow the interface interact with the Grid Network. These calls are based on the use of another referenced project package (GridwayBridge). All the needed parameters for different calls are taken both from the introduced information through the interface and from the temporary files created or updated at a new job throwing in the local host to keep some job data

It's recommended to use the visual GUI editor provided by the Visual Studio suite in order to alter the look of the application's GUI.

B.3 Gridway Bridge

The application is composed

Introduction

It is the library that allows the communication between plugin_excel_grid_runner and the server gridway (gwd). It is developed with C# language.

This library is completely separated from the other components of plugin excel gridway, which use it as black box.

It offers support to plugin_excel_gridway's component (plugin excel grid runner) through functions that emulate gridway's commands, in order to the plugin excel grid runner abstracts of the aspects of connection with the server. The library GridWayBridge is in charge of communication with the remote server to send the works, to monitorize them and to transmit the files needed between the machine client and the server gridway.

The connections and transferences of files are made under SSH protocol, doing use of the library SharpSSH (A Secure Shell library for .NET)

It is important to remember that this library always considers that we are working with arrays of works, and never simple works.

General description

- | |
|--|
| <ol style="list-style-type: none">1. Configuration data2. Operational commandos with grid3. Internal support utilities |
|--|

1. Data configuration

Set of methods that allow to manage the configuration data necessary to be able to operate with the server gridway (gwd).

Configuration data that handles the library:

- - Server host (ip or machine name)
- - Linux user

- - Linux user's password
- - Proxy's key
- - Console's prompt

Since finally connection is under SSH, it is necessary to know the user's name, and the password it will use to connect to the server.

To use Gridway it is essential that the user of grid has some credentials. It is necessary to have the proxy's key that allows to active the credentials to be able to make use of grid.

Finally, it is also necessary to know the prompt of the server's console, since the dialogue with the server is based on the console writing and the delay of answer, which detects when the prompt arrives (ex: \$>).

The methods that the library implements to manipulate these data are:

```
bool esta_inicializadoSSH()
```

It indicates if the library knows the data access SSH to the server machine (user and Linux password, host and prompt).

```
bool esta_inicializadoProxy()
```

It indicates if the library knows the key that activates proxy.

```
void inicializaSSH(string prompt_, string host_,string usuario_, string clave_)
```

It indicates to the library the data access SSH to the server machine (user and Linux password, host and Prom.).


```
void inicializaProxy(string clave_)
```

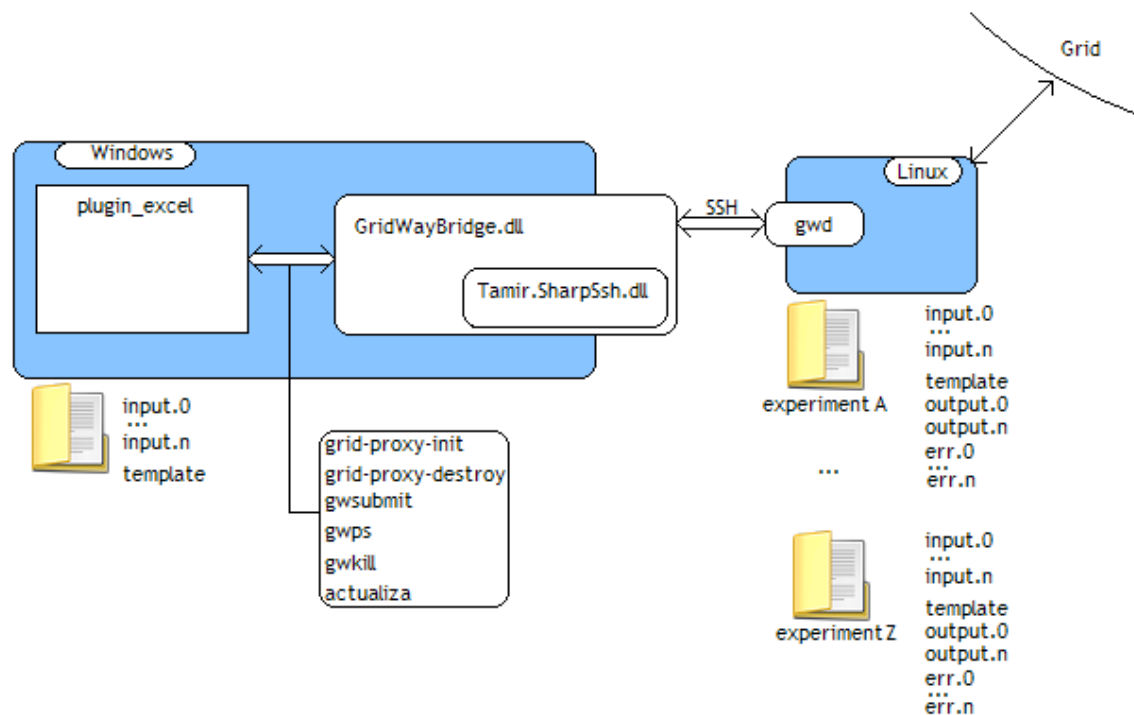
It indicates to the library the key to active proxy.

2. Grid commands to operate with grid

They provide interaction with the server for the shipment, monitorization and obtaining of results of works.

We remember that it is important not to forget that this library always considers that we are operating with arrays of works, and never simple works.

Operation scheme:



It is possible to observe that each array job will have a folder where its information will be stored. This folder will be a necessary datum for the shipment and the reception of results.

The implemented methods in the library are:

```
string grid_proxy_init()
```

Emulation of grid_proxy_init Globus command. The function gives back a chain with activity information. So that it is able to complete its task, the key of proxy must be initialized previously.

```
string grid_proxy_destroy()
```

Emulation of grid_proxy_init Globus command. The function gives back a chain with activity information.

```
int gwsubmit(string carpeta, string template, string ruta, int num_tareas, string ruta_ejecutable)
```

Emulation of gwsubmit GridWay command.

The function gives back the generated array jobs identifier.

```
public static int[] gwps(int[] AIDS)
```

Emulation of gwps GridWay command.

The function receives an Array with identifiers of the Array jobs that is wanted to verify.

An Array of the same dimensions that the one of entrance is returned, where the position i-ésima of this array, will have a whole number that indicates the number of subtasks finished of the “array job” of the position i-ésima of the entrance vector, considering itself finished those with state “it donates” or “fail”, and not

finished is given back in another case.

```
public static bool gwkill(int AID, string carpeta)
```

Emulation of gwkill GridWay command.

It receives the identifier of “Array job” that is wanted to cancel.

It gives back true if it has been cancelled correctly and false on the contrary.

3. Utilidades de apoyo internas de la librería

```
bool enviaFichero(string carpeta, string fichero)
```

It sends a file from the local folder of work of the program to the server and it places it in the folder indicated like parameter. It gives back certain if the shipment were satisfactory, and false in another case.

```
bool recogeFichero(string carpeta, string fichero, string rutadestino)
```

It gathers a file of the indicated remote folder as parameter (experiment folder) and it places it in the premises with the route specified by the rutadestino parameter. It gives back certain if the shipment were satisfactory, and false in another case.

Apéndice C - HowTo Español

C.1 Plugin_excel

La aplicación se compone de tres paquetes o módulos.

El primer modulo es “plugin-excel”.

Este paquete está basado en la construcción de una extensión para Microsoft Excel mediante Microsoft Visual Basic .NET.

Estas extensiones se llaman extensiones COM, y Microsoft Office 2000 y las versiones posteriores cuentan con una arquitectura de diseño uniforme para construir este tipo de aplicaciones que extienden, mejoran y controlan las aplicaciones de Office.

La manera de construir este tipo de extensiones se describe paso a paso en este artículo, así como las subrutinas más importantes de este módulo.

Una extensión COM es un servidor COM en proceso, o una librería de enlace dinámico ActiveX (DLL), que implementa la interfaz IDTextensibility2 como se describe en la librería de tipos del diseñador de extensiones de Microsoft (Msaddndr.dll). Todas las extensiones COM heredan de esta interfaz y deben implementar cada uno de sus cinco métodos.

De esta forma, este paquete contiene la clase Connect.vb, que implementa estos métodos, que son:

OnConnection

El evento se lanza cuando el plugin se conecta. Si tiene éxito en la ejecución, se lanza el plugin. Pero en caso de error, se devuelve un mensaje informando del mismo, y la aplicación inmediatamente libera su referencia al plugin, y se destruye el objeto.

Recibe los cuatro parámetros siguientes:

- Application (Aplicación) - Una referencia al objeto de la aplicación.
- ConnectMode (Modo de conexión) - Una constante que especifica como se conecta el plugin.
- AddInInst (Instancia del AddIn) - Una referencia a la extensión COM que referencia a dicha extensión en la colección de extensiones COM de la aplicación.

- Custom - Un array de valores de tipo Variant que contienen datos definidos por el usuario.

Usamos los parámetros Application y AddInInst para inicializar la aplicación Excel.

OnDisconnection

El evento se lanza cuando se desconecta el plugin y justo antes de liberar la memoria. El plugin realiza una limpieza de recursos en este evento.

Recibe los dos parámetros siguientes:

- RemoveMode (Modo de eliminación) - Una constante que especifica como se desconecta el plugin.
- Custom - Un array de valores de tipo Variant que contienen datos definidos por el usuario.

OnAddInsUpdate

El evento se lanza cuando el conjunto de los plugins COM registrados cambia. En otras palabras, cuando un plugin COM se instala o se elimina de la aplicación.

OnStartupComplete

El evento es llamado cuando el plugin se conecta durante el lanzamiento. Su funcionalidad es crear un nuevo botón en el menú “Herramientas” de Microsoft Excel con los parámetros iniciales apropiados (como el nombre y la visibilidad), asegurándose que el botón del menú no existía previamente.

OnBeginShutdown

El evento se lanza cuando el host desconecta el plugin durante el cierre.

O_cmd_Click

El método es lanzado cuando el usuario pulsa el botón de la barra de herramientas de Microsoft Excel, “plugin excel grid”. Su funcionalidad es conectar el plugin con el formulario que está incluido en el siguiente paquete, plugin_excel_grid_runner, y llama a este paquete para cargar el formulario.

Para construir un plugin como éste, se debe crear un nuevo Proyecto en Microsoft Visual Studio .NET, y en “Otros proyectos” seleccionar Proyectos de Extensibilidad, y después seleccionar la plantilla de Extensiones Compartidas. El siguiente paso es darle un nombre al

plugin, seleccionar crear el plugin usando Visual Basic, y después seleccionar la aplicación de Microsoft para la que se debería construir el plugin. Y siguiendo estos pasos, se forma la clase Connect detallada previamente.

C.2 Plugin_excel_grid_runner

Dos son las principales funciones que aporta este paquete:

- Mostrar un interfaz amigable y útil al usuario.
- Implementar los métodos de llamada que permiten al interfaz la posibilidad de trabajar con las funciones propias de las redes grid mediante la ayuda del paquete GridwayBridge.

Este paquete es una simple aplicación de Windows compuesta de clases propias de Visual Basic (VBA). Cabe reseñar el detalle de que contiene dos referencias especiales a otros componentes:

- Excel Interop Library (Microsoft.Office.interop.Excel), que permite la interacción de el paquete y sus métodos con la aplicación Microsoft Excel.
- El paquete GridwayBridge, de tal manera que nos provee de funcionalidad a la hora de trabajar con redes grid.

A continuación plantearemos la manera en que cada clase cumple su cometido

GSettings

En esta clase tienen cabida aquellos métodos necesarios para el tratamiento de la información que vamos a utilizar a la hora de crear un nuevo trabajo. En concreto maneja tanto información acerca del nuevo trabajo como de los detalles del usuario que está ejecutando la aplicación en ese momento.

Esta clase utiliza ‘serialización’, que es un proceso propio de Visual Basic que permite convertir un objeto en una secuencia lineal de bits. Esta herramienta resulta de gran utilidad porque permite un notable ahorro de espacio y una mayor eficiencia respecto al tiempo de acceso y almacenamiento de la memoria cuando se crean nuevos trabajos.

Esta clase implementa los métodos get y set relativos a los campos de datos.

GTrabajos

De la misma forma que GSettings, esta es otra clase que utiliza ‘serialización’ y aporta una manera de tratar mejor la información con la que se trabaja. En este caso la información versa en torno a los detalles específicos de un trabajo y se almacenará en el equipo local desde el que se lance la ejecución, de tal manera que pueda recuperarse cada vez que se soliciten detalles acerca del trabajo.

Mappings

Gracias a esta clase se puede generar un fichero ‘.dat’ en el que se almacenan todos los datos relativos a la configuración que se han gestionado a través de GSettings y Gtrabajos

mainClass

Esta clase gestiona el lanzamiento de la aplicación cuando se le llama desde Microsoft Excel a través de la opción destinada a ello.

frmGClient.vb

Se trata de una clase de Visual Basic común que utiliza el paquete ‘System.Windows.Form’ con la finalidad de generar el interfaz gráfico. Todos los métodos y componentes que permiten a esta clase crear un GUI se encuentran en ese paquete. Esta clase además utilizará GSettings y Gtrabajos para gestionar la información que se está manipulando tras el interfaz.

El interfaz de usuario cuenta con tres secciones diferentes que atañen diversos aspectos:

- **Settings:** En este apartado se reflejan los parámetros del usuario. Estos parámetros son necesarios para la creación de nuevos trabajos o para la consulta de los detalles de alguno ya creado previamente. Estos datos pueden guardarse en el equipo local gracias a la clase Mappings.
- **Job Submit:** Es en esta sección donde el usuario puede especificar y enviar nuevos trabajos, definiendo para ellos un nombre, el archivo a ejecutar y los parámetros de entrada para ese archivo tomados directamente desde las celdas e Microsoft Excel. Se puede elegir entre la opción de enviar el ejecutable desde el equipo local que esté utilizando el usuario o la de utilizar otro tomado directamente de una máquina remota.
- **Onitoring:** Aquí se muestra la información acerca de los trabajos que han sido enviados (y cuyos resultados no se han recogido aun). Todos los datos acerca de estos se recogen en un DataTable que se muestra al usuario: job id, nombre, estado, etc. Además cabe la posibilidad de centrarse aun más en un determinado trabajo, recogiendo aun más detalles de este como por ejemplo el número total de tareas de

las que se compone o el número de estas que ya se han realizado. Para conseguir mostrar todos estos datos se toma la información del módulo GridwayBridge y de un archivo creado con la ayuda de las clases GTrabajos y Mappings, el cual se actualiza cada vez que se realiza alguna acción sobre un trabajo.

La actualización de una hoja de Excel con los resultados de un trabajo finalizado es otra de las funciones de este apartado, tomándolos de la máquina remota en los que están alojados a través del paquete GridwayBridge de nuevo. Además se puede cancelar un trabajo que ya ha sido creado mientras no se haya solicitado la recuperación de los resultados hasta el equipo local.

Esta clase contiene por añadidura aquellas funciones que permiten al interfaz trabajar directamente con las redes grid; como se ha reseñado en puntos anteriores, para poder realizar estas llamadas el módulo se basa en el paquete GridwayBridge. Todos los parámetros que resultan necesarios a la hora de llevar a cabo las distintas llamadas son tomados tanto de la información que se refiere a una consulta concreta introducida mediante el interfaz como de los archivos temporales creados y actualizados con el fin de mantener de forma permanente cierta información acerca de los trabajos en el equipo local.

Se recomienda el uso del editor visual para GUI's provisto por la suite de Visual Studio si se desea modificar el aspecto del GUI de la aplicación.

C.3 Plugin ExcelGridway

Introducción

Es la librería que permite la comunicación entre plugin_excel_grid_runner y el servidor gridway (gwd). Está desarrollada en lenguaje C#.

Esta librería está completamente desacoplada de los demás componentes de Plugin Excel Gridway, que lo utilizan como caja negra.

Ofrece soporte, al componente plugin_excel_grid_runner de Plugin Excel Gridway, mediante funciones que emulan comandos de gridway, de manera que plugin_excel_grid_runner se abstraer de los aspectos de conexión con el servidor. La librería GridWayBridge se ocupa de comunicarse con el servidor remoto para lanzar los trabajos, monitorizarlos y transmitir los ficheros necesarios entre la máquina cliente y el servidor gridway.

Las conexiones y transferencias de ficheros se realizan bajo el protocolo SSH, haciendo uso de la librería SharpSSH (A Secure Shell library for .NET)

Es importante tener en cuenta que esta librería considera siempre que se opera con arrays de trabajos, y nunca trabajos simples.

Funcionamiento general consta de tres partes diferenciadas:

- | |
|-----------------------------------|
| 1. Datos de configuración |
| 2. Comandos de operación con grid |
| 3. Utilidades de apoyo internas |

1. Datos de configuración

Conjunto de métodos que permiten controlar los datos de configuración necesarios para poder operar con el servidor gridway (gwd).

Datos de configuración que maneja la librería:

- Host servidor (IP o nombre de máquina)
- Usuario Linux
- Clave de usuario Linux
- Clave de Proxy
- Prompt de consola

Puesto que finalmente la conexión es bajo SSH, es necesario conocer el nombre de usuario y contraseña que se utilizará para conectarse al servidor.

Para la utilización de Gridway es imprescindible que el usuario que utiliza el grid tenga unas credenciales. Se necesita la clave del proxy que permita activar las credenciales para poder hacer uso del grid.

Por último necesario conocer el prompt de la consola del servidor, puesto que el diálogo con el servidor está basado en la escritura en consola y la espera de respuesta, que se detecta cuando llega el prompt (ej: \$>).

Los métodos que implementa la librería para manipular estos datos son:

```
bool esta_inicializadoSSH()
```

Indica si la librería conoce los datos de acceso SSH a la máquina servidor (usuario y contraseña Linux, host y prompt).

```
bool esta_inicializadoProxy()
```

Indica si la librería conoce la contraseña de activación de proxy.

```
void inicializaSSH(string prompt_, string host_,string usuario_, string clave_)
```

Indica a la librería los datos de acceso SSH a la máquina servidor (usuario y contraseña Linux, host y prompt).

```
void inicializaProxy(string clave_)
```

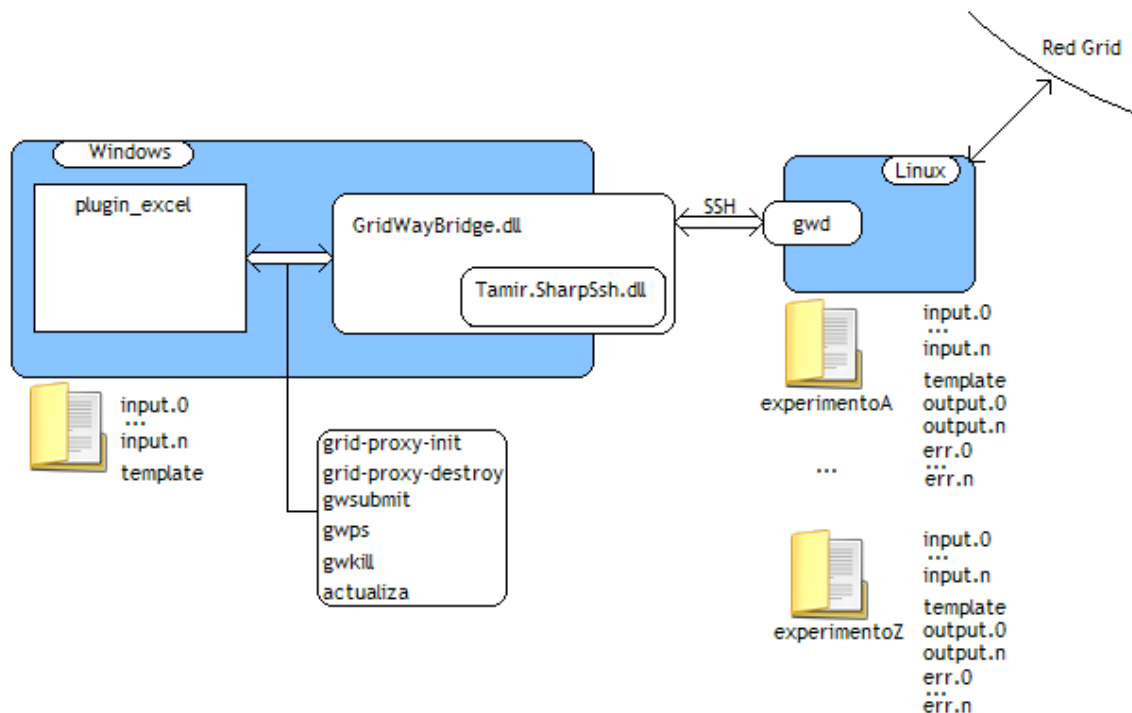
Indica a la librería la contraseña de activación de proxy.

2. Comandos de operación con grid

Proporcionan interacción con el servidor para el envío, monitorización y obtención de resultados de trabajos.

Recordamos que es importante tener en cuenta que esta librería considera siempre que se opera con arrays de trabajos, y nunca trabajos simples.

Esquema de operación:



Se puede observar que cada array de jobs tendrá una carpeta donde se almacenará su información. Esta carpeta será un dato necesario para el envío y la recepción de resultados.

Los métodos implementados en la librería son:

```
string grid_proxy_init()
```

Emulación del comando Globus `grid_proxy_init`. La función devuelve una cadena con información de actividad. Para que consiga completar su tarea debe estar la clave de proxy inicializada previamente.

```
string grid_proxy_destroy()
```

Emulación del comando Globus `grid_proxy_init`. La función devuelve una cadena con información de actividad.

```
int gwsubmit(string carpeta, string template, string ruta, int num_tareas, string  
ruta_ejecutable)
```

Emulación del comando GridWay gwsubmit.

La función devuelve el identificador de array de jobs generado.

```
public static int[] gwps(int[] AIDS)
```

Emulación del comando GridWay gwps.

La función recibe un array con identificadores de los array jobs que se quiere comprobar.

Se devuelve un array de las mismas dimensiones que el de entrada, donde la posición *i*-ésima de este array, tendrá un número entero que indica el número de subtareas terminadas del "array job" de la posición *i*-ésima del vector de entrada, considerándose terminadas aquellas con estado "done" o "fail", y no terminadas en otro caso.

```
public static bool gwkill(int AID, string carpeta)
```

Emulación del comando GridWay gwkill.

Recibe el identificador de "array job" que se quiere cancelar.

Devuelve true si se ha cancelado correctamente y false en caso contrario.

3. Utilidades de apoyo internas de la librería

```
bool enviaFichero(string carpeta, string fichero)
```

Envía un fichero desde la carpeta local de trabajo del programa hasta el servidor y lo coloca en la carpeta indicada como parámetro. Devuelve cierto si el envío fue satisfactorio, y falso en otro caso.

```
bool recogeFichero(string carpeta, string fichero, string rutadestino)
```

Recoge un fichero de la carpeta remota indicada como parámetro (carpeta de experimento) y lo coloca en local con la ruta especificada por el parámetro rutadestino. Devuelve cierto si el envío fue satisfactorio, y falso en otro caso.

Apéndice D - Bibliografía

PÁGINAS WEB CONSULTADAS:

Grid Computing por José Manuel Molina
para el *Centro de Difusión de Tecnologías ETSIT-UPM*
[http://www.ceditec.etsit.upm.es/grid_computing.php]

Introducción a la computación en Grid por Rubén Santiago Montero e Ignacio Martín Llorente
para CURSO SUPERIOR DE ADMINISTRACIÓN, EXPLOTACIÓN Y PROGRAMACIÓN DE SISTEMAS
GRID de la Escuela Complutense de Verano
[<http://asds.dacya.ucm.es/cursogrid.php>]

Globus Toolkit 4. Web Service Resource Framework por Rubén Santiago Montero e Ignacio
Martín Llorente para CURSO SUPERIOR DE ADMINISTRACIÓN, EXPLOTACIÓN Y PROGRAMACIÓN
DE SISTEMAS GRID de la Escuela Complutense de Verano
[<http://asds.dacya.ucm.es/cursogrid.php>]

¿What is the Globus Toolkit? por Globus Toolkit para Globus Project
[<http://www.globus.org/toolkit/>]

El metaplanificador Gridway
[<http://www.gridway.org>]

ExcelGrid: A .NET Plug-in for Outsourcing Excel Spreadsheet Workload to Enterprise and
Global Grids por Krishna Nadiminti, Yi-Feng Chiu, Nick Teoh, Akshay Luther, Srikumar
Venugopal, Rajkumar Buyya1
[<http://www.gridbus.org/excelgrid/>]

Platform Symphony
[<http://www2.platform.com/products/Symphony/>]

Excel Connector
[<http://www.gridsystems.com/>]

Las hojas de cálculo: evolución y mercado
[<http://www.utdallas.edu/~liebowitz/book/sheets/sheet.html>]

ActiveSheets: Super-Computing with Spreadsheets por David Abramson, Paul Roe, Lew Kotler,
Dinelli Mather
[<http://www.utdallas.edu/~liebowitz/book/sheets/sheet.html>]

ActiveSheets: Super-Computing with Spreadsheets
[<http://www.csse.monash.edu.au/~davida/papers/ActiveSheets.pdf>]

SharpSSH - A Secure Shell (SSH) library for .NET
[<http://www.tamirgal.com/home/dev.aspx?Item=SharpSsh>]
